

Schreiber, David

135252

From: Ramirez, Delia
Sent: Thursday, October 14, 2004 1:10 PM
To: Schreiber, David
Subject: 10/041018

Hi,

I would like to request the following searches:

1. seq id 1 in the nucleic acid databases (commercial and interference)
2. seq id 361 in the nucleic acid databases (commercial and interference)
3. seq id 383 in the protein databases (commercial and interference)
4. seq id 22 in the protein databases (commercial and interference)
5. alignment of seq id 22 and seq id 1
6. alignment of seq id 383 and seq id 361

Thank you,

Delia M. Ramirez, Ph.D.
Patent Examiner
Recombinant Enzymes-Art Unit 1652
USPTO
400 Dulany Street, Remsen Bldg., 2D74, Mail room 2C70
Alexandria, VA 22314
(571) 272-0938
delia.ramirez@uspto.gov

3

SEARCH REQUEST FORM**Scientific and Technical Information Center**

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: <u>D. Schreiber</u>	NA Sequence (#) <u>4</u>	STN _____
Searcher Phone #: <u>272-2526</u>	AA Sequence (#) <u>4</u>	Dialog _____
Searcher Location: <u>Remsen E01 #61</u>	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: _____	Bibliographic _____	Dr.Link _____
Date Completed: <u>10/18</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: <u>15</u>	Fulltext _____	Sequence Systems <u>CompuGen</u>
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: <u>27</u>	Other _____	Other (specify) _____

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 03:30:51 ; Search time 292.161 Seconds
(without alignments)
1543.991 Million cell updates/sec

Title: US-10-041-018-383

Perfect score: 4113

Sequence: 1 MNLSCIASPLITKSNRPAA.....TVKDIYNPLVLNENRQR 784

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt.02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4113	100.0	784	2 Q9XEI0	Q9xei0 stevia reba
2	4026	97.9	784	2 Q9XEI9	Q9xeh9 stevia reba
3	2855.5	69.4	788	2 Q9FXV8	Q9fxv8 lactuca sat
4	2101.5	51.1	785	2 Q9PRX5	Q9prx5 cucumis sat
5	2023.5	49.2	789	2 Q93548	Q93548 cucurbita m
6	2021	49.1	785	2 Q64952	Q64952 arabidopsis
7	1530	37.8	683	2 Q7XLE0	Q7xle0 oryza sativ
8	1530	37.2	821	2 Q6Z5V6	Q6z5v6 oryza sativ
9	1530	37.2	821	2 BAD17270	Bad17270 oryza sat
10	1530	37.2	830	2 Q85282	Q85282 oryza sativ
11	1530	37.2	802	2 Q8V934	Q8v934 oryza sativ
12	1525.5	37.1	802	2 BAD17672	Bad17672 oryza sat
13	1525.5	37.1	802	2 Q7XX63	Q7xx63 oryza sativ
14	1319	32.1	770	2 Q7XLD9	Q7xld9 oryza sativ
15	1312	31.9	913	2 Q7XLE2	Q7xle2 oryza sativ
16	1276	31.0	730	2 Q7XLE2	Q7xle2 oryza sativ
17	1144.5	27.8	673	2 Q6Z5J0	Q6z5j0 oryza sativ
18	1144.5	27.8	673	2 BAD17276	Bad17276 oryza sat
19	1102	26.8	590	2 Q84ZW8	Q84zw8 zea mays (m
20	1075	26.1	873	2 Q947C4	Q947c4 ginkgo bilo
21	1075	26.1	873	2 AAS89668	Aas89668 ginkgo bi
22	1062.5	25.8	868	2 Q38710	Q38710 abies grand
23	1051.5	25.6	853	2 Q94FW1	Q94fw1 abies grand
24	1017.5	24.7	862	1 TASY_TAXBA	Tasy_taxba taxus bac
25	1017.5	24.7	862	2 Q8TBY0	Q8tby0 taxus bacca
26	1017.5	24.7	862	2 AAR02861	Aar02861 taxus bac
27	1016.5	24.7	862	2 Q6SA60	Q6sa60 taxus x med
28	1016.5	24.7	862	2 AAS18603	Aas18603 taxus x m
29	1013.5	24.6	862	1 TASY_TAXBR	Tasy_taxbr taxus brevi
30	990.5	24.1	862	1 TASY_TAXCH	Tasy_taxch taxus chine
31	965	23.5	782	2 Q9SAU6	Q9sau6 abies grand

32	965	23.5	817	2	O81086	O81086 abies grand
33	962	23.4	816	2	Q94FW2	Q94fw2 abies grand
34	925	22.5	458	2	Q9XEI5	Q9xei5 zea mays (m
35	876.5	21.3	871	2	Q92PN5	Q92pn5 clarkia con
36	863	21.0	870	2	Q96376	Q96376 clarkia bre
37	853	20.7	877	2	Q93YV0	Q93yv0 arabidopsis
38	837.5	20.4	813	2	Q92PN4	Q92pn4 clarkia bre
39	837.5	20.4	878	2	Q84U03	Q84u03 arabidopsis
40	813.5	19.8	870	2	O22733	O22733 arabidopsis
41	760	18.5	829	2	Q9MAX2	Q9max2 croton subl
42	732	17.8	787	2	O22667	O22667 stevia reba
43	725.5	17.6	801	2	O04408	O04408 pisum sativ
44	723.5	17.6	479	2	Q6VTF8	Q6vtf8 oryza sativ
45	723.5	17.6	479	2	BAD17771	Bad17771 oryza sat

ALIGNMENTS

RESULT 1

Q9XEI0 PRELIMINARY; PRT; 784 AA.

ID Q9XEI0; AC Q9XEI0; DT 01-NOV-1999 (TREMELrel. 12, Created)

DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE kaurene synthase.

GN Name=KS22-1;

OS Stevia rebaudiana (Stevia).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;

OC campanulids; Asterales; Asteraceae; Eupatorieae; Eupatoriaceae; Stevia.

OX NCBI_taxid=55670;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Leaf;

RX MEDLINE=99435880; PubMed=10504563;

RA Richman A.S., Gijzen M., Starratt A.N., Yang Z., Brandle J.B.;

RT "Diterpene synthesis in Stevia rebaudiana: recruitment and up-

RL Plant J. 19:411-421(1999).

DR EMBL; AF097311; AAD34295.1; --

DR HSSP; Q40577; 1HX9.

DR GO; GO:0016829; F:lyase activity; IEA.

DR GO; GO:0008152; P:metabolism; IEA.

DR InterPro; IPR005630; Terpene synth C.

DR InterPro; IPR008949; Terpenoid synth.

DR InterPro; IPR008930; Terp cyc toroid.

DR InterPro; IPR001906; Terp synth-like.

DR Pfam; PF01397; Terpene synth_1.

DR Pfam; PF03936; Terpene synth_C; 1.

SQ SEQUENCE 784 AA; 89400 MW; 284BD3972DC09051 CRC64;

Query Match 100.0%; Score 4113; DB 2; Length 784;

Best Local Similarity 100.0%; Pred. No. 4.1e-234;

Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MNLSCIASPLITKSNRPAA	SAIHTASTSHGGQTNP	TNLIIDTTKRIQKFN	VSIV 60
Db	1	MNLSCIASPLITKSNRPAA	SAIHTASTSHGGQTNP	TNLIIDTTKRIQKFN	VSIV 60
Qy	61	SSYDTAWAVPSPNSPKS	CFPECLNWLNNQ	LDGSGVLNHNHPL	KDLSST 120
Db	61	SSYDTAWAVPSPNSPKS	CFPECLNWLNNQ	LDGSGVLNHNHPL	KDLSST 120
Qy	121	LACTVALKRNVEDQ	INKGLSFIESNL	ASATEKSPSPIGFI	IFPGILEYAKNLDINL 180
Db	121	LACTVALKRNVEDQ	INKGLSFIESNL	ASATEKSPSPIGFI	IFPGILEYAKNLDINL 180
Qy	181	LSKQTFDPLMLHRELE	QKCHSNEMDGYL	AYISEGLGNLYD	WNWVKYQMKNGSVFNSP 240
Db	181	LSKQTFDPLMLHRELE	QKCHSNEMDGYL	AYISEGLGNLYD	WNWVKYQMKNGSVFNSP 240

QY 241 SATAAAFINHQPGCLNVLNSLLDKFGNAVPTVYPHDLFIRLSWMDTIERLGISHFRVE 300
 Db 241 SATAAAFINHQPGCLNVLNSLLDKFGNAVPTVYPHDLFIRLSWMDTIERLGISHFRVE 300
 QY 301 IKNVLDYTRCWRVDEQIFMDVVTTCALAFRLRLRINGEYVSPDLAEITNELALKDEYAA 360
 Db 301 IKNVLDYTRCWRVDEQIFMDVVTTCALAFRLRLRINGEYVSPDLAEITNELALKDEYAA 360
 QY 361 LETYHSHILYQEDLSSGKQILKSADFLKEIISTDSNRSLKLIHKEVENALKPPIINTGLE 420
 Db 361 LETYHSHILYQEDLSSGKQILKSADFLKEIISTDSNRSLKLIHKEVENALKPPIINTGLE 420
 QY 421 RINTRENIQIYNVDNTRILKTYHSSNI SNTDYRLA VEDFTYCQSIYREELKGLERWV 480
 Db 421 RINTRENIQIYNVDNTRILKTYHSSNI SNTDYRLA VEDFTYCQSIYREELKGLERWV 480
 QY 481 ENKLDQKFAOKTAYCYFSAATLSSPELSDARISWAKNGILTTVVDDFFDIGGTIDEL 540
 Db 481 ENKLDQKFAOKTAYCYFSAATLSSPELSDARISWAKNGILTTVVDDFFDIGGTIDEL 540
 QY 541 TNLIOQVEKNWVDVDCCEHVRILFLALKDAICWIGDEAFKQWQARDVTSHVIOTWLEL 600
 Db 541 TNLIOQVEKNWVDVDCCEHVRILFLALKDAICWIGDEAFKQWQARDVTSHVIOTWLEL 600
 QY 601 MNSMLREAIWTRDAYPTLNEYMENAYVSPALGPVVKPAIYFVGPKLSEIIVSSSYHNL 660
 Db 601 MNSMLREAIWTRDAYPTLNEYMENAYVSPALGPVVKPAIYFVGPKLSEIIVSSSYHNL 660
 QY 661 FKLMSTQGRLLNDIHSFKREFKEGKLNALVALHLSNGESGKVEEVEEMMMIKNKKREL 720
 Db 661 FKLMSTQGRLLNDIHSFKREFKEGKLNALVALHLSNGESGKVEEVEEMMMIKNKKREL 720
 QY 721 MKLIFENGSIIVPRACKDAFWNMCHVLPFFYANDDGTGTNTILDTVKDIINPLVLVYEN 780
 Db 721 MKLIFENGSIIVPRACKDAFWNMCHVLPFFYANDDGTGTNTILDTVKDIINPLVLVYEN 780
 QY 781 EQOR 784
 Db 781 EQOR 784

RESULT 2

Q9XEH9 Q9XEH9 PRELIMINARY; PRT; 784 AA.
 AC Q9XEH9;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Kaurene synthase.
 GN Names:KSI-1;
 OS Stevia rebaudiana (Stevia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC campanulids; Asterales; Asteraceae; Eupatorieae; Eupatoriaceae; Stevia.
 OX NCBI_TaxID=55670;
 [1]
 RN
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leaf;
 RX MEDLINE=99435880; PubMed=10504563;
 RA Richman A.S., Gijzen M., Starratt A.N., Yang Z., Brandle J.E.;
 RT "Pterpene synthesis in Stevia rebaudiana: recruitment and up-
 regulation of key enzymes from the gibberellin biosynthetic pathway.";
 RL Plant J. 19:411-421(1999).
 DR EMBL; AF097310; AAD34294.1; -.
 DR HSSP; O81192; INLZ.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR005630; Terpene synth C.
 DR InterPro; IPR008949; Terpenoid synth.
 DR InterPro; IPR008930; Terp_cyc_Teroid.
 DR InterPro; IPR001906; Terp_synth-like.
 DR Pfam; PF01397; Terpene synth; 1.
 DR Pfam; PF03936; Terpene synth_C; 1.

SQ SEQUENCE 784 AA; 89401 MW; D83B95DE21C0B35C CRC64;
 Query Match 97.9%; Score 4026; DB 2; Length 784;
 Best Local Similarity 98.0%; Pred. No. 5.6e-229;
 Matches 768; Conservative 8; Mismatches 8; Indels 0; Gaps 0;
 QY 1 MNLSLCIASPLITKSNRPAALSAIHTASTSHGQNTPTNLIIDTTKERTOKQKFNVEISV 60
 Db 1 MNLSLCIASPLITKSNRPAALSAIHTASTSHGQNTPTNLIIDTTKERTOKQKFNVEISV 60
 QY 61 SSIYTAWAMVSPSPKSPCPPECLNMLINNQLDNGSWGLVNHHTHHNHPILKDSLSST 120
 Db 61 SSIYTAWAMVSPSPKSPCPPECLNMLINNQLDNGSWGLVNHHTHHNHPILKDSLSST 120
 QY 121 LACTIVALKRWVNGEDQINKGLSFISNLSASATEKQSPSPIGFDIIPPGLEVAKNLDINL 180
 Db 121 LACTIVALKRWVNGEDQINKGLSFISNLSASATEKQSPSPIGFDIIPPGLEVAKNLDINL 180
 QY 181 LSKQTDFFSLMLHRELEQKCHSNEMDGYLAVISEGLNLYDMNMVVKYQMKNGSVFNSP 240
 Db 181 LSKQTDFFSLMLHRELEQKCHSNEMDGYLAVISEGLNLYDMNMVVKYQMKNGSVFNSP 240
 QY 241 SATAAAFINHQPGCLNVLNSLLDKFGNAVPTVYPHDLFIRLSWMDTIERLGISHFRVE 300
 Db 241 SATAAAFINHQPGCLNVLNSLLDKFGNAVPTVYPHDLFIRLSWMDTIERLGISHFRVE 300
 QY 301 IKNVLDYTRCWRVDEQIFMDVVTTCALAFRLRLRINGEYVSPDLAEITNELALKDEYAA 360
 Db 301 IKNVLDYTRCWRVDEQIFMDVVTTCALAFRLRLRINGEYVSPDLAEITNELALKDEYAA 360
 QY 361 LETYHSHILYQEDLSSGKQILKSADFLKEIISTDSNRSLKLIHKEVENALKPPIINTGLE 420
 Db 361 LETYHSHILYQEDLSSGKQILKSADFLKEIISTDSNRSLKLIHKEVENALKPPIINTGLE 420
 QY 421 RINTRENIQIYNVDNTRILKTYHSSNI SNTDYRLA VEDFTYCQSIYREELKGLERWV 480
 Db 421 RINTRENIQIYNVDNTRILKTYHSSNI SNTDYRLA VEDFTYCQSIYREELKGLERWV 480
 QY 481 ENKLDQKFAOKTAYCYFSAATLSSPELSDARISWAKNGILTTVVDDFFDIGGTIDEL 540
 Db 481 ENKLDQKFAOKTAYCYFSAATLSSPELSDARISWAKNGILTTVVDDFFDIGGTIDEL 540
 QY 541 TNLIOQVEKNWVDVDCCEHVRILFLALKDAICWIGDEAFKQWQARDVTSHVIOTWLEL 600
 Db 541 TNLIOQVEKNWVDVDCCEHVRILFLALKDAICWIGDEAFKQWQARDVTSHVIOTWLEL 600
 QY 601 MNSMLREAIWTRDAYPTLNEYMENAYVSPALGPVVKPAIYFVGPKLSEIIVSSSYHNL 660
 Db 601 MNSMLREAIWTRDAYPTLNEYMENAYVSPALGPVVKPAIYFVGPKLSEIIVSSSYHNL 660
 QY 661 FKLMSTQGRLLNDIHSFKREFKEGKLNALVALHLSNGESGKVEEVEEMMMIKNKKREL 720
 Db 661 FKLMSTQGRLLNDIHSFKREFKEGKLNALVALHLSNGESGKVEEVEEMMMIKNKKREL 720
 QY 721 MKLIFENGSIIVPRACKDAFWNMCHVLPFFYANDDGTGTNTILDTVKDIINPLVLVYEN 780
 Db 721 MKLIFENGSIIVPRACKDAFWNMCHVLPFFYANDDGTGTNTILDTVKDIINPLVLVYEN 780
 QY 781 EQOR 784
 Db 781 EQOR 784
 RESULT 3
 Q9FXV8 Q9FXV8 PRELIMINARY; PRT; 788 AA.
 ID Q9FXV8;
 AC Q9FXV8;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Ent-kaurene synthase No1.
 GN Name=LKSI1;
 OS Lactuca sativa (Garden lettuce).

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
Campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae;
Lactuca.
OX NCBI TaxID=4236;
RN [1]_

RP SEQUENCE FROM N.A.
RC STRAIN-Grand Rapids;
KA Toyomasu T., Mitsuhashi W., Kamiya Y.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB031205; BABI2441.1; -.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPRO05630; Terpeno synth C.
DR InterPro; IPRO008949; Terpenoid synth.
DR InterPro; IPRO008930; Terp cyc toroid.
DR InterPro; IPRO01906; Terp synth-like.
DR Pfam; PF01397; Terpene synth; I.
DR Pfam; PF03936; Terpene synth C; 1.
SQ SEQUENCE 788 AA; 8961 MW; 0AA211132810CE7 CRC64;

Query Match 69.4%; Score 2855.5; DB 2; Length 788;
Best Local Similarity 67.8%; Pred. No. 6.4e-160;
Matches 541; Conservative 102; Mismatches 116; Indels 39; Gaps

QY 1 MNLSICIASPLLTGKN-----RPAALSAIHASTSHGGQNPTMLIIDT 44
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 45 TKERIQOFKNEVTSVSYSYDAWAMVPSPNSPKSPCPPECLNLMNLNNQLNDGSGLNVNH 104
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 57 TKERIRKLFENNVEVSYSYDAWAMVPSPNSPKSPCFPCDCLNWLNDQLDDGSGLL-- 114
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 105 THNNHPLLKDLSSTLACIALVKLRNVGEDQINKGLSFIBSNLASATEKSQPPIGFDI 164
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 115 --PHQSPLIKDTLSSTLACVLAKRWNVKQDKINKGLHYIESNTASVTDKNQASPFGFDI 172
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 165 IFPELLLEVAKNLDINLLSKOTDFSLMLHKRELEOKRCHSNEMDGVIAYISEGLNLYDWN 224
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 173 IFPGMLEYAKDLDKLPINQTHLSVMHLHERELRRCHSGNGREAYLAYISEGLNMDNW 232
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 225 MVKYQMKGNSVFNSPATAAFINHQNPGLINTYLLDKFGNAPVTYPDHDLFIRISM 284
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 233 VMVKYQMKGSLFNPSPATASVLHHQNAAGLHYLTLLDKFGNAPVTVPIDILYRISM 292
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 285 VDIETRLGI SHHFVEITKNULDETYRCWVERDEQIFMDVVTCALAFKRLRINGEVSFDP 344
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 293 VDTLERLGIKRHFVEIQNVLDITYRCWQGDVQIFMDVVTCALAFRVLRNGEVSFSDP 352
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 345 LAELITN-----ELAKDYEALEYTHASHIIYLQEDLSGKKILKASADFLKEIISTDS 396
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 353 LAKITEGDYMSPEKPFKVYTSLVYKASQIIYQEELAFREONLYTS-----YLPSS 405
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 397 NRISKLHKVEVALKFPINTGLERINTRRIQIYNVDNRTRILKTTHSSNISMTYLR 456
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 406 NKLSNYILKEVDALKFPFNGSLERMSTRRIEHYNLHTRILKTTYSSSNSINKDYLYL 465
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 457 AVDEFTYCOSTYREELKGERWWVENKLDQIKFAKQTATCYFSVAATLSPELSDARIS 516
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 466 AVODEMECOSIYCEEKDLERWVENKLDKLFARQRTATCYFSAASFSLSPDISDARIS 525
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 517 WAKNGILTTVVDDFFDIGTIIDELTNLIQCVEKNVNDVKDCSEHVRIILFLAKDKAIWC 576
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 526 WAKSSILLTTVIDFFDVGGSDDELWNFVHIEKNVNVENDCSEEVGVLFLAKDKAVCW 585
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 577 IGDEAFKWQRADVTSHVIOTWLEIMNSMLREAITWRDAYVPTLNEMYENAYVFSALGPV 636
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 586 IGDKAFXIQERNITSHTVETWLDLVKSLREAIWAKDGSIPTINEYMENGVSFALGPV 645
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 637 KPATYFVGPKLSEIVSVSSVHNLFKLMSTQGRLLINDHSFKREFKCKGLNAVALLHNSG 696
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 646 LPTLYIFGVKLSSEVWGSEYHKLYFMYSTQGRLMNDIHSFKREKAKGKLNALVMSDG 705
Db |||:::||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 697 ESKGVEEVEVEMMMKMKRKLMLKLFPENGSIVPRACKDAFNNWCHVNFNFYANDDG 756
 Db KSGSVEEVEEEMKILTSQRKEMKVLVLETKGSVVRVCKDFVNNCNVNLFYATDDG 765
 QY 757 FTGNLTDTVDKIIYNPL 774
 Db FTGNALDWDVKEIIYEPV 783

RESULT 4
 Q9FRX5 PRELIMINARY; PRT; 785 AA.

AC Q9FRX5;
 DT 01-MAR-2001 (TremBLrel. 16, Created)
 DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
 DE Ent-kaurene synthase.
 GN Name=CskS1;
 OS Cucumis sativus (Cucumber).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
 OX NCBI_taxid=3659;
 RN [1]

SEQUENCE FROM N.A.
 RP Shira K., Fujino K., Masuda K.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB045310; BAB19275.1; -
 DR GO; GO:0016829; P:lyase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR001330; Prenyltrans.
 DR InterPro; IPR005630; Terpene synth C.
 DR InterPro; IPR008949; Terpenoid synth.
 DR InterPro; IPR008930; Terp_cycteroid.
 DR InterPro; IPR001906; Terp_synth-like.
 DR Pfam; PF00432; Prenyltrans; 1.
 DR Pfam; PF01397; Terpene synth; 1.
 DR Pfam; PF03936; Terpene synth C; 1.
 SQ SEQUENCE 785 AA; 88706 MW; 8EE26D97D33PB350 CRC64;

Query Match 51.1%; Score 2101.5; DB 2; Length 785;
 Best Local Similarity 50.9%; Pred. No. 1.9e-115;
 Matches 413; Conservative 136; Mismatches 194; Indels 69; Gaps 10

QY 1 MNLSTCIASPLLTNSRPAALSAHTASTSH-----GGQTNPTNLIIDTKERIQKQ 52
 Db 1 MNL-----RPTNLGCF-TASSASLPGLDVGTKTGTGARFEETKERIKKL 47

QY 53 FKNVEISVSYDTAWAVMPSNPSPKSPCFPECLNWLINQNDGSGVLNHNHNHPL 112
 Db FKNVELSISAYDTAWAVMPSNLSNKLPLFPECINWLHDHQPDSWGIL-----HDHQL 102

QY 113 -LKDSLSTLACIVALKRWNVGDEQINGKLSFIESNLASATEKQSPSIGFDIIPPGLE 171
 Db VMKATLSTLACVLTLEKRWIDIGDHMSKALSFIKSNIASATDENQKSPSGVDIIPPGMIE 162

QY 172 YAKNLDINLLSKQTFSLMLHKLEOKRCHSNEMDG--YLAYISEGIGNLYDNWVKKY 229
 Db YAKDLNINLPLASNVDVALVQKLEELRSCSNSEGGKAYLAYVSEGTGLQDNEMVMRY 222

QY 230 QMKNGSVNFSPTAAAFINHQNPGLNLYNLSLDFKGNVPTWYPDHLFIRLSMVDTIE 289
 Db QRKNGSLFSPSTTAVAFMRNDGCFNLYRLSVLQKPHSSVFAIYPLDIYARLHWVDSLQ 282

QY 290 RLGISHHFRVEIKNVLDETRCWVERDEQIFMDVVTICALPRLLRINGEYVSPDPLABIT 349
 Db KLGIDGHFKDEIRSVLDETTSCWNGQENIFLDASTCAMAFLMLRVEGYDVSDDLQTFQS 342

QY 350 NEL-----ALKDEYAALETVYHASHLYQED-----LSSGKQILKS 384
 Db EGLFNSCLGHLKDFASLFLFKASQIIIVYDPEILENINSWTSRFLNHHLSGG----- 396

QY 385 ADFLKEIITSDNRLSKLIHKEVENALKFPINTGLERINTFRNLTQLYVNDTRILKITYH 444


```

Db 500 CYFSGAATLSPSELSARISWAKGGVLTWVDDFDVGGSGKELENLHLVEKWDLNGVP 559
QY 557 DCSSEHVRILFLAKDAICWIGDEAFKQWQARDVTSVHVIQTWLELNMMLREAIWTDAV 616
Db 560 EYSSEHEIIFSVLRDITLETGDKAFTYQGRNVTHHVIKVLDDLKSMLEAEWSDKST 619
QY 617 PTLNEMENAYVSFALGPVTKPAIYFVGPGLSEIEVESSEYHNLFKLMSTQGRLLNDIHS 676
Db 620 PSELYMENAYVSFALGPVLPATYLLIGPPEKTVDSHQYNQYKLVSTGMGLLNDIQ 679
QY 677 FKREFKEGKNAVALHLSNGESGKVEEVVEVMMMKIKRKEMLKLIPEENGSIIVPRAC 736
Db 680 FKRESAEGKNAVALHLSNGESGKVEEVVEVMMMKIKRKEMLKLIPEENGSIIVPRAC 739
QY 737 KDAFWNMCHVIAFFYANDGFTNTILDTVKDIYNPLVLNVNEN 780
Db 740 KEAFLKMSVLYNLFRKDKDFTSNDLSLVKSVIYEFVLSQES 783

RESULT 8
Q7XLE0 PRELIMINARY; PRT; 683 AA.
AC Q7XLE0;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DS UNB0070C17.8 protein.
GN Names:OSUNB0070C17.8;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,
RA Weng Q., Zhang L., Lu Y., Mu Y., Zhang L.S., Yu Z., Fan D.,
RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,
RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
RA Chen J., Kang H., Shao C., Sun X., Shao X., Zhang X., Zhang W.,
RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
RA Han B.;
RT "Sequence and analysis of rice chromosome 4.";
RL Nature 420:316-320(2002).
DR EMBL; AL731610; CAB05201.3; -.
DR Gramene; Q7XLE0; -.
DR GO; GO:0016829; F-lyase activity; IEA.
DR GO; GO:0008152; P-metabolism; IEA.
DR InterPro; IPR005630; Terpene synthase C.
DR InterPro; IPR001906; Terp synthase-like.
DR Pfam; PF01397; Terpene synthase 1.
DR Pfam; PF03936; Terpene synthase C.
SQ SEQUENCE 683 AA; 77861 MW; B18A2387456CAB84 CRC64;

Query Match 37.8%; Score 1556.5; DB 2; Length 683;
Best Local Similarity 43.5%; Pred. No. 2.3e-83;
Matches 323; Conservative 114; Mismatches 202; Indels 103; Gaps 12;

QY 48 RIQKQFNKVBISVSYDTAWAVVPSPKSPCFPCPLNWLNNQNLGSGWLVNHTN 107
Db 19 RTDQQLTSLTSLDVTAWAVVPLGSRHQHCFPCQVENVILNQDQDSWG----TRG 74
QY 108 HNHPLKXDSLSTLACTIVALKRWNVGSDQNKGLSFIESMLASATEKSPSGFDIIPP 167
Db 75 FGAVTRDVLSTLACVALKRWNVGQEHIRRGILDFGRNFSTAMDRQIAAPVGFNITFP 134
QY 168 GLLEYAKNLDINLLSKOTDFSLMHLKELQKR---CHSNEMDGYLAVISGGLNLYDWN 224
Db 135 GMLSLGMDLDFPVRQTDVRLHLHLREIELERAGDHSYGRKAYMAVTEGLNLEWD 194

```

```

QY 225 MVKTYQMKNQSVFNSPSATAAAFIHQNPGLNLSLIDKFGNAVPTVYPHDLFIRLSM 284
Db 195 EIMWFQKNGSFNFCSTTAAITLVHNDKALQYLNCLL----- 233
QY 285 VDTIERLGHSHFRVEIKNVLDVTRCWVERBQIFMDVVVTCALAPRLIRINGEYVSP-- 342
Db 234 -----VSSWLERDEEIMLDITTCAMAPRLRMNGYHYVSSVE 270
QY 343 -DPLAEITN-----ELAKDEYAALETYHSHLYQEDLSSGKQLKLSADFLKEIISTDS 396
Db 271 LSPVBAASSFRSLQYLNKSLIELYKASKWKY----- 305
QY 397 NRLSKLIHKEVENALKFPINTGLERINTRNIQLNVNDTRILKTTYHSSNISTDYRL 456
Db 306 -----ALKFPYVTLDRLDHKNRIERFADKDSQMLKTEYLLPH-ANQDILAL 351
QY 457 AVEDFVTCOSIYREBELKGLERWVVENKLDQKFAROKTAYCYPVSAATLSSPELSDAIS 516
Db 352 AVEDFSSQSISYQDELNLYLECWVKDKLPQFARQKLTTCYKLSAAATIPPELSARIA 411
QY 517 WAKNGILTVDVDDFDIGGTIDELTNLIQCVKKNVDVDDKCCSEHVRILFLAKDAICW 576
Db 412 WAKNGVLTVDVDDFDLGGSKKELENLIALVEKND-CHQEEFYSEQVRIVFSIYTVNQ 470
QY 577 IGDEAFKQWQARDVTSVHVIQTWLELNMMLREAIWTDAVPTLNEMENAYVSFALGPV 636
Db 471 LGAKASALQGRDVTKHLEIWLCLMESWMTAEWQRTKYVPTMEETMANAVVSFALGPV 530
QY 637 KPAIYFVGPGLSEIEVESSEYHNLFKLMSTQGRLLNDIHSFKREFKEGKNAVALHLSNG 696
Db 531 LPTLYFVGPGLQEDVVRDHEYNELFRIMSTCGRLNDSQGFERESLEGKLSVSL-LVHH 589
QY 697 ESGKVE-EWVEREMMMIKRKLMLKLIPEENGSIIVPRACKDAFNMWCHLVNFFVAND 755
Db 590 SGGISIDAKWKAQKSIITSRNLRILVIGEQA-VPRCKOLFWMCKVIMFVFSRTD 648
QY 756 GFTG-NTILDVTKDIYNPLVL 776
Db 649 GFSSPKEMVSANNAVVKPEPLK 670

RESULT 9
Q6Z5J6 PRELIMINARY; PRT; 821 AA.
AC Q6Z5J6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Putative ent-kaurane synthase B.
GN Names:P0689H05.20;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005114; BAD17270.1; -.
DR InterPro; IPR005630; Terpene synthase C.
DR InterPro; IPR008949; Terpene synthase.
DR InterPro; IPR008930; Terpene synthase.
DR InterPro; IPR001906; Terpene synthase-like.
DR Pfam; PF01397; Terpene synthase 1.
DR Pfam; PF03936; Terpene synthase C.
SQ SEQUENCE 821 AA; 92376 MW; 52557D85FE875B11 CRC64;

Query Match 37.2%; Score 1530; DB 2; Length 821;
Best Local Similarity 41.2%; Pred. No. 1.1e-81;
Matches 332; Conservative 151; Mismatches 259; Indels 64; Gaps 16;

```


DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative diterpene cyclase.
 GN Name=OsDTC1;
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=14675427;
 RA Cho E., Okada A., Kemmoku H., Otomo K., Toyomasu T., Mitsuhashi W.,
 RA Sassa T., Yajima A., Yabuta G., Mori K., Oikawa H., Toshihima H.,
 RA Shibuya N., Nojiri H., Omori T., Nishiyama M., Yamane H.;
 RT "Molecular cloning and characterization of a cDNA encoding ent-cassa-
 RT 12,15-diene synthase, a putative diterpene phytoalexin biosynthetic
 RT enzyme, from suspension-cultured rice cells treated with a chitin
 RT elicitor.";
 RL Plant J. 37:1-8 (2004).
 DR EMBL; AB089272; BAC56714.1; -.
 DR Gramene; Q85282; -.
 DR GO; GO:0016829; F1ylase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR005630; Terpene synthase.
 DR InterPro; IPR008949; Terpene synthase.
 DR InterPro; IPR008930; Terpene synthase.
 DR InterPro; IPR001906; Terpene synthase.
 DR Pfam; PF01397; Terpene synthase; 1.
 DR Pfam; PF03936; Terpene synthase; 1.
 DR SEQUENCE 830 AA; 92091 MW; 1F23A5A43B703835 CRC64;
 Query Match 37.2%; Score 1530; DB 2; Length 830;
 Best Local Similarity 43.0%; Pred. No. 1.1e-81;
 Matches 324; Conservative 122; Mismatches 265; Indels 42; Gaps 13;
 QY 49 IQQFKNVLSVSSYDTAWVAVPSPNSPKSPCEPCLNWLNNQNDGSGVLNHTNH 108
 DB 92 IRKQGVLSVSSYDTAWVAVPSPNSPKSPCEPCLNWLNNQNDGSGVLNHTNH 151
 QY 109 NHPELLKDSLSSTLACIALKRNWVGDQINKGSLFIESNLASATEKSQSPGFIIFPG 168
 DB 152 N----KDVLSSTLACIALKRNWVGDQINKGSLFIESNLASATEKSQSPGFIIFPG 207
 QY 169 LLEYAKNLDINLSKOTDPSMLHKELEOKR---CHSNEMDGYLAIYSEGLNLYDNWM 225
 DB 208 MLTLANGSGLEVVRQNDIDSLNHLREMKIQREAGNHSRGRKAYMAYLAEGFGLNLEWDE 267
 QY 226 VKKYQMKNGSVFNSPSATAAFINHQNPCLNLYSLDKFGNAVPTVPHDLFIRLSMV 285
 DB 268 IMFFQRNGSLFNCPSSTAGALANYHDDKALQVLSLVNKFQGVVPTLYPLNYCQLSMV 327
 QY 286 DTLERLGI SHHFRVEIKNVLDIETRCWVERDEQIFMDVVTTCALAFRLRLINGEVSDDPL 345
 DB 328 DALENMGI SQYFASIKSILDMTYSYSLWLGKDEIMLDVTTCAVAFRLRLWNGYDVSSDEL 387
 QY 346 AETNELA-----LKDEYAALETVHAS-HILYQEDLSGSKQILKS-----ADPLKE 390
 DB 388 SHVAGSGFRDLSQGYLNDKRSVLEVKYTKSHISENDL-----ILDSIGSGSGGLLKE 442
 QY 391 IISTDSNRLSKLHKEVENALKEPINTGLERINTRRNIOLYVNDNTRILKT---YHSSN 447
 DB 443 MLS--SNCKGTGPEEELFALKYFYSTLERLVRKRNVLFDKGSQMLKTCMPVHDSQ 500
 QY 448 ISNTDYLRLAVEDFYTCQSYREELKGLERWVVENKLDQKFAKQKTAICYFVAATLSS 507
 DB 501 ----DFLAVDVDFCISQSNQNELNYLESWKDNRDLQHFARQKITYCYVLSGAATFR 556
 QY 508 PELSDARLSWAKNGILTVVDDFDIGGTTDELNTLIQCVKKNVDVDDKCCSEHVLIF 567
 DB 557 PEMGYARTSWARTAWLTAVIDDLDLFDVGLGQEQENLLALMEKWEPEGEDEYISDVKIVF 616

QY 568 LALKDAICWIGDEAFKQWQARDVTSHTVQIOTWLELMSMLREAIWTRDAYVPTINEMENAY 627
 DB 617 QALYNTVNEIGAKASALQCHDVTKLVDVWLHVRCMKVEAWQSRSHLPTEFYMESGM 676
 QY 628 VSFALGPVIVKPAIYFVGPKLSEIEVESSEYHMLFKLMSTQGRLLNDIHSHKREFFKGL- 686
 DB 677 VSLGQGCCTVMSALFLIGEXKLPBGIVLEBYDELFLMGTCGRLLNDIRGERESDGKMT 736
 QY 687 NAVAL--HLSNGESGKVEEVEVMMMTKNRKKELMKLIFENGSIIVPRACKDAFNNWC 744
 DB 737 NGVSLVHASGSGMS--VDEAKTEVMKRIDARRKLLSLVSEQEGPIPRPCKQLFWKMC 794
 QY 745 HVLNPFYANDGFTG-NTILDVTWKDIIYNPLVL 776
 DB 795 KILHLFYOTDGFSPKEMVSAVDVINEPLQL 827
 RESULT 12
 Q6YV94 PRELIMINARY; PRT; 802 AA.
 AC Q6YV94;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Putative ent-kaurene synthase.
 GN Name=OSUNBA0008E01.23;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sasaki I., Matsumoto T., Katayose Y.;
 RA Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP005835; BAD17672.1; -.
 DR InterPro; IPR005630; Terpene synthase.
 DR InterPro; IPR008949; Terpene synthase.
 DR InterPro; IPR008930; Terpene synthase.
 DR InterPro; IPR001906; Terpene synthase.
 DR Pfam; PF01397; Terpene synthase; 1.
 DR Pfam; PF03936; Terpene synthase; 1.
 DR SEQUENCE 802 AA; 89665 MW; 1D45CB1FC46E42AC CRC64;
 Query Match 37.1%; Score 1525.5; DB 2; Length 802;
 Best Local Similarity 43.0%; Pred. No. 1.9e-81;
 Matches 321; Conservative 122; Mismatches 273; Indels 31; Gaps 11;
 QY 49 IQQFKNVLSVSSYDTAWVAVPSPNSPKSPCEPCLNWLNNQNDGSGVLNHTNH 108
 DB 65 IRKQGVLSVSSYDTAWVAVPSPNSPKSPCEPCLNWLNNQNDGSGVLNHTNH 124
 QY 109 NHPELLKDSLSSTLACIALKRNWVGDQINKGSLFIESNLASATEKSQSPGFIIFPG 168
 DB 125 N----KDVLSSTLACIALKRNWVGDQINKGSLFIESNLASATEKSQSPGFIIFPG 180
 QY 169 LLEYAKNLDINLSKOTDPSMLHKELEOKR---CHSNEMDGYLAIYSEGLNLYDNWM 225
 DB 181 MLTLANGSGLEVVRQNDIDSLNHLREMKIQREAGNHSRGRKAYMAYLAEGFGLNLEWDE 240
 QY 226 VKKYQMKNGSVFNSPSATAAFINHQNPCLNLYSLDKFGNAVPTVPHDLFIRLSMV 285
 DB 241 IMFFQRNGSLFNCPSSTAGALANYHDDKALQVLSLVNKFQGVVPTLYPLNYCQLSMV 300
 QY 286 DTLERLGI SHHFRVEIKNVLDIETRCWVERDEQIFMDVVTTCALAFRLRLINGEVSDDPL 345
 DB 301 DALENMGI SQYFASIKSILDMTYSYSLWLGKDEIMLDVTTCAVAFRLRLWNGYDVSSDEL 360
 QY 346 AETNELA-----LKDEYAALETVHAS-HILYQEDLSGSKQILKSADFLKEIISTDS 396
 DB 361 SHVAGSGFRDLSQGYLNDKRSVLEVKYTKSHISENDL-----ILDSIGSGSGGLLKEMLC--S 418
 QY 397 NRLSKLHKEVENALKEPINTGLERINTRRNIOLYVNDNTRILKT---YHSSNINSDY 453

```

Db 419 NGKGTGPRREIEFALKYFYSTLERLVHRKNIIVLPDAKGSQMLKTECMPVHDSQ----DF 474
QY 454 LRLAVDEFYTCOSIYREELKGLERWVENKLDQKFAKQKTAICYPSVAATLSSPESLSDA 513
Db 475 LALAVDDFCISQSNYNELNYSWVKDNLQDLHFAKQKITCYLSGAATFRPEMGYA 534
QY 514 RISWAKNGILITVDDFFDGGTIDELTNLIOCVEKNVVDKDCGSEHVRILFLALAKDA 573
Db 535 RTSWARTAWLTAVIDDLDVGGLEQOEENLLALMEKWEPEGDEYYSEYSDVKIVFOALYNT 594
QY 574 ICWIGDEAFKQARVDTSHVITQWLELMNSMLREAIWTRDAYVPTLNEYMENAYVSFALG 633
Db 595 VNEIGAKASALQGHVDVTKYLVWLVHVRCKMVEAEWQSQHLPTFEETMESGMVSLGOG 654
QY 634 PIVKPAIYFVGPKLSEIEVESSEYHNLFKLMSTQGRLLNDIHSFKREFKGL-NAVAL- 691
Db 655 CTVMASALFLIGEKLPFGIVLEBEYDFELFRLMGTQGRLLNDIRGIBREESDGRWTVGSLL 714
QY 692 -HLSNGESGKVEEVVEEMMMIKNKRKELMKLIIFEENGSIIVPRACKDAFNMCHVLNFF 750
Db 715 VHASGGSMS--VDEAKTEVMKRIDASRRKLLSLVSEQEGPIPRPCKQLFWKCKILHLF 772
QY 751 YANDGFTG-NILDTVKDIINPLVL 776
Db 773 YYQTDGFSPEKEMVSAVDVINEPLQL 799

RESULT 13
BADI7672 PRELIMINARY; PRT; 802 AA.
AC BADI7672;
DT 10-MAY-2004 (TREMELrel. 27, Created)
DT 10-MAY-2004 (TREMELrel. 27, Last sequence update)
DE Putative ent-kaurene synthase.
GN OSUNBA0008E01.23.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza; Oryza sativa.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, BAC
clone:OSUNBA0008E01."
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005835; BADI7672.1; -.
SQ SEQUENCE 802 AA; 89665 MW; 1D45CB1FC46E42AC CRC64;

Query Match 37.1%; Score 1525.5; DB 2; Length 802;
Best Local Similarity 43.0%; Pred. NO. 1.9e-81;
Matches 321; Conservative 122; Mismatches 273; Indels 31; Gaps 11;

QY 49 IQQFKNVEISVSYDTAWAVPSPKSPCEPCNLNLLNQLNDGSLVNHHTNH 108
Db 65 IRKQLQGVLESPSYDTAWAVPVRGSHNPSFPQCVWILENQWDGDSIDGSTA 124
QY 109 NHPLLKSLSTLACTIVALKRWNVGEDQINKGLSFIESNLASATEKSPSPGFDIIPFG 168
Db 125 N----KDVLSSTLACVLALNKNVGRHIRRLGSLFGRNFSIAMDDQAVAPIGGITFPA 180
QY 169 LLEYAKNLIDNLKQDTFSLMLHKELEQKR----CHSNEMDGYLAYISEGLNLYDNWM 225
Db 181 MLTLANGSGLEVPVRQNDIDSINHLREMKIOREAGNHSRGRKAYMAYLAEGFGNLEWDE 240
QY 226 VKYQMKNGSVNPSATAAFINHQPCNLNLYSLDKGNVAVPTVYPHDLFIRLSMV 285
Db 241 IMMFQRKNGSLFNCPSSFAGALANVHDDKALQYLQSLVKNFGDVGVVPTLPLNIYCQLSMV 300
QY 286 DTIBELGISHFVRVEIKNVLDVTCWVERDEQIFMDVVTICALAFRLIRINGEVSPDPL 345

```

```

Db 301 DALENWGISQYFASEIKSILDNTYSSWIGDKBEIMLDVTTCAWAFRLRMNGDYVSSDEL 360
QY 346 ABITNELA-----LKDEYAALETIHAS-HILYQEDLSSGKQILKSAADFLKEIISTDS 396
Db 361 SHVAGASGFRDLSQGVNDRKSVLEVYKTSKHSISENDLILDSIGSWSSGLKEMLC--S 418
QY 397 NELSILIKHEVENALKPPIINTGLEIRNTRNTQLNVNDNTRILKT---TYHSSNISNTDY 453
Db 419 NGKGTGPRREIEFALKYFYSTLERLVHRKNIIVLPDAKGSQMLKTECMPVHDSQ----DF 474
QY 454 LRLAVDEFYTCOSIYREELKGLERWVENKLDQKFAKQKTAICYPSVAATLSSPESLSDA 513
Db 475 LALAVDDFCISQSNYNELNYSWVKDNLQDLHFAKQKITCYLSGAATFRPEMGYA 534
QY 514 RISWAKNGILITVDDFFDGGTIDELTNLIOCVEKNVVDKDCGSEHVRILFLALAKDA 573
Db 535 RTSWARTAWLTAVIDDLDVGGLEQOEENLLALMEKWEPEGDEYYSEYSDVKIVFOALYNT 594
QY 574 ICWIGDEAFKQARVDTSHVITQWLELMNSMLREAIWTRDAYVPTLNEYMENAYVSFALG 633
Db 595 VNEIGAKASALQGHVDVTKYLVWLVHVRCKMVEAEWQSQHLPTFEETMESGMVSLGOG 654
QY 634 PIVKPAIYFVGPKLSEIEVESSEYHNLFKLMSTQGRLLNDIHSFKREFKGL-NAVAL- 691
Db 655 CTVMASALFLIGEKLPFGIVLEBEYDFELFRLMGTQGRLLNDIRGIBREESDGRWTVGSLL 714
QY 692 -HLSNGESGKVEEVVEEMMMIKNKRKELMKLIIFEENGSIIVPRACKDAFNMCHVLNFF 750
Db 715 VHASGGSMS--VDEAKTEVMKRIDASRRKLLSLVSEQEGPIPRPCKQLFWKCKILHLF 772
QY 751 YANDGFTG-NILDTVKDIINPLVL 776
Db 773 YYQTDGFSPEKEMVSAVDVINEPLQL 799

RESULT 14
QYXX63 PRELIMINARY; PRT; 770 AA.
AC QYXX63;
DT 01-OCT-2003 (TREMELrel. 25, Created)
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE OSUNBA0052P16.14 proteain.
GN Name=OSUNBA0052P16.14;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12447439;
RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,
RA Wang Q., Zhang L., Lu Y., Wu J., Lu Y., Zhang L.S., Yu Z., Fan D.,
RA Liu X., Zhu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,
RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
RA Han B.;
RT "Sequence and analysis of rice chromosome 4.";
RL Nature 420:316-320(2002).
DR EMBL; AL662936; CAD39717.1; -.
DR Gramene; Q7XX63; -.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR005630; Terpene synthase C.
DR InterPro; IPR001906; Terp synth-like.
DR Pfam; PF01397; Terpene synthase 1.
DR Pfam; PF03936; Terpene synthase 2.

```

SQ SEQUENCE 770 AA; 86960 MW; 6CFAB6855D904EE1 CRC64;
Query Match 32.1%; Score 1319; DB 2; Length 770;
Best Local Similarity 37.7%; Pred. No. 2.7e-69;
Matches 288; Conservative 146; Mismatches 223; Indels 106; Gaps 16;
QY 32 GGOTNPTLIIDTTERIKQKFNVEISVSSYDVAWVMPFNSP--KSPCFPECLNWL 89
DB 89 GENTSQINM--GWEARIILHLENPEFLPSSDYIAWVMPFEGTHLOAPCEPCEVWI 145
QY 90 INNOLMDGSGVLVNHHTNHNHLLKDSLSSTLACTIVALKRWVNGEDQINKGLSFTIESNLA 149
DB 146 LQNHSGNSGWG---NEFDSSAKDILLSTLACIIALEKWNVSGEIRRLHFIKFNFS 201
QY 150 SATEKSPQSPIDFIIFPGLELVAKNLDINLLSKQTFDPSMLHKRELEQKCHSNE--M 206
DB 202 IVIDDQIAPIGNTFFPAWNLAIKNGLEFFASEISIDQILHLRDMELKRLAGDSGLK 261
QY 207 DGLAYISGL-ONLYDNWVKYQMKNGSVFNSPATAAFINHQNGCLNYLNSLLDK 265
DB 262 EAYFAIAGLESWVDNVEVMKFGKNGSLFNSPATAAALVHYRDDKALGYLYSVNKK 321
QY 266 FGNVPTVPHDLFIRLSWDTTIERIGSHHFVEIKNVLDIETYRCWVERDEQIFMDVVT 325
DB 322 FGGEVPTVPLNIFSQLSMVDTLVNIIGISRHFSDDIKRIIDKTYILWSQDEEWMLDLPT 381
QY 326 CALAPILLIRINGEYSPDPLAETN-----ELAKDEYAALFTYVHASHILYQEDLS 377
DB 382 CAMAFLLRMNGYGVSSDDLSSHVAEASTFNSVEGYLDLDTKSLLEYKASKV-----SLSE 437
QY 378 GKQILK-----SADFLKEIISTDSNRILKIHKEVENALKFPINTGLERINTERNIQLYN 432
DB 438 NEPILEKMGCSGLLKEKLCSDDIRGTPIL-REVEVALKFPFYATLEPLDHKNLENFD 496
QY 433 VDNTRILKTYHSNTSNTDYLRLAVEDFTYQCSYREELKGLERWVVENKLDQLKPARQ 492
DB 497 ARAYQIKTKNMPCHV-NEDLLAALAEEDFSQOSTYQNETQHLERW-----DQ----- 543
QY 493 KTAICYFSYAATLSSPELSDARTSWAKNGLLITTVDDFFDIDGTTIDELTNLIQCVKWNV 552
DB 544 -----YHKV-----EFY----- 550
QY 553 DVBDKCCSBEHRLFLALKDAICWIGDEAFKQARDVTSHVITWLELMNSMLREAIWTR 612
DB 551 -----SENVKAVFALYSTVNLGAMASAVQNRQVTKYNVESWLDYLSLATDAEWQR 603
QY 613 DAVPTLNEYMNAYVSFALGPVKPAIYFVGPKLSEIEVESSEYHNLFKLMSTQGLLN 672
DB 604 SKYVPTMEETKNSIVTFALGPITLIALYFMQNLWEDIVKNAEYDFELFRIMNTQGLQN 663
QY 673 DIHSFKREEKGLNAVALHLSNGESGKVEEVEEMMMKKNRKMELMKLIPEENGSIY 732
DB 664 DIOSFERCKDGKINSVSLVDSKQVMSVEEAKENESISSCRRELLRLVVRDQ-VI 722
QY 733 PRACKDAFNMWCHLVNPFYDNDGFTG-NTILDVTKDIYNPL 774
DB 723 PKSCKEFWNLKXTSHVYFYSQADGFSPKEMGMGAVGVIPEPL 765

RESULT 15

Q7XLD9 PRELIMINARY; PRT; 913 AA.
AC Q7XLD9
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE OSJNBa0070C17.9 protein.
GN Name=OSJNBa0070C17.9;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;

RN SEQUENCE FROM N.A.
RP PubMed=12447439;
RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,
RA Wang Q., Zhang L., Lu Y., Mu Y., Zhang L.S., Yu Z., Fan D.,
RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,
RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
RA Lan B., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
RA Han B.;
RT "Sequence and analysis of rice chromosome 4.";
RL Nature 420:316-320(2002).
DR EMBL; AL731610; CAE05202.3; --
DR Gramene; Q7XLD9; --
DR GO; GO:0016829; Filyase activity; IEA.
DR GO; GO:0008152; Pimetalolism; IEA.
DR InterPro; IPR005630; Terpene synth C.
DR InterPro; IPR001906; Terp synth-like.
DR Pfam; PF01397; Terpene synth; 1.
DR Pfam; PF03936; Terpene synth C; 1.
SQ SEQUENCE 913 AA; 99830 MW; 3FF5D613C92C6C7F CRC64;

Query Match 31.9%; Score 1312; DB 2; Length 913;
Best Local Similarity 38.4%; Pred. No. 9e-69;
Matches 289; Conservative 142; Mismatches 232; Indels 90; Gaps 20;
QY 70 MUPS-PNSPKSPCFPECLNWLNNQNDGWS--GLVNHHTNHP--LLKDSLSSTLACI 124
DB 1 MVDALRGGGGPRFPQCVAMIQRNQRGDGSRHAAAHQOLGSSPRIVTERDLSSTLACV 60
QY 125 VALKRWVGEQDQINKGLSFTIESNLASATEKSQSPITGPDII-PPGLELVAKNLDINLLS- 182
DB 61 LALARWDAGEHVRRLGQIFGRNNSVAMDDQTAAPASGVSVAAMLRWAMENGLEVPV 120
QY 183 -----KQTFSLMLHKRELEQKCHSNEMDGYLAYISEGLNLYDNWVKYQMKNGSVF 237
DB 121 SQADVDRDRDAGVI-----CHGRTE-YTAYVSEGLGNTQNNWVNMKFORKNGSLF 169
QY 238 NSPSATAAFINQNGCLNLYNSLIDKFGNAVPTVYVPHDLFIRLSWDTTIERIGTSHHF 297
DB 170 NSPTTAAALVNHVNDKALQYLDLDDKFGSAVPAAPANIQOLYQVNDVLEKMGISRH 229
QY 298 RVKIKVLDIETYRCWVERDEQIFMDVVTALAFRLRLRINGEYVSPDPLAET-----NE 351
DB 230 VGIKSIILDMTYSQWQDEIVLDMQTCGMAFRMLRMNGYDVSDELSEFSEPSPHNS 289
QY 352 LA--LKDEYAALFTYVHASHILYQEDLSGKQILKSDAFLKEIISTDSNRILSKLIHKEVEN 409
DB 290 LQGLVNDTRSLLELHKASKV-----STAEKEVEY 318
QY 410 ALKFPINTGLERINTERNIQLYNVDNTRILKTYHSNINSDTYLRLAVEDFTYQCSYIR 469
DB 319 ALEFPFTILDRDLRDKRNIHFDTSSQMLETAYLPCH-SNEIMALGVDFDSSSQIFQ 377
QY 470 EELKGLERWVVENKLDQLKPARQKAYCFVSAATLSSPELSDARTSWAKNGIITVVD 529
DB 378 EELQOLNSWKESRLDQLQAPAKLDYFSAATITFPELSDVRLWAKNGVITVVD 437
QY 530 FFDIGGTIDELTNLIQCVKWNVDKCCSEHVRILFLAKDAICWIGDEAFKQARDV 589
DB 438 FFDVGGSGKEELNVALVEKWDKNDKTEYVSEQVEIVFSAITYSTNQLGSMASVQGRDV 497
QY 590 TSHVIQTLWLELMNSMLREAIWT--RDAYVPTLNEYME-NAYVVSFALGPI-----VK 637
DB 498 TKHLE-----IVSIISDILWNVFFSVRTDKNLIVASVNLCLIVARFAEYVDEGRVEAE 552
QY 638 PAIFYVGPKLSEIEVESSEYHNLFKLMSTQGLNDIHSFKFEKGLNAVALHLSNGE 697
DB 553 PHCILLGPKMPSVIRSQCESELFRMLSKGRLINDVQSVYEREGSQGLKNSVSL-LALHS 611

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 08:05:06 ; Search time 51.1457 Seconds
(without alignments)
1474.884 Million cell updates/sec

Title: US-10-041-018-383
Perfect score: 4113
Sequence: 1 MNLSLCIASPLTKSNRPA.....TVKDIYNPLVLVNEBQR 784

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79.*
1: Piri.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2023.5	49.2	789	T09672	ent-kaurene synthase
2	2021	49.1	785	G96825	hypothetical prote
3	2021	49.1	785	T52059	ent-kaurene synthase
4	813.5	19.8	870	A96637	hypothetical prote
5	725.5	17.6	801	T06783	ent-kaurene synthase
6	709	17.2	802	D85035	ent-kaurene synthase
7	657.5	16.0	823	T02959	kaurene synthase A
8	446.5	10.9	591	H84633	probable limonene
9	431	10.5	600	F71434	probable limonene
10	426	10.4	559	T08174	sesquiterpene cycl
11	413	10.0	599	A48863	limonene cyclase -
12	406	9.9	1024	T71434	probable limonene
13	399	9.7	550	T03714	5-epi-aristolochene
14	398.5	9.7	520	A56118	ent-kaurene synthase
15	373	9.1	554	S68366	(+)-delta-cadinene
16	367	8.9	554	S68385	germacrene C synth
17	357.5	8.7	548	T06285	hypothetical prote
18	354	8.6	548	T06266	hypothetical prote
19	354	8.6	632	E96723	hypothetical prote
20	326	7.9	608	G86443	probable terpene s
21	321	7.8	598	H86460	hypothetical prote
22	290	7.1	535	T06285	hypothetical prote
23	288	7.0	530	G96588	hypothetical prote
24	278	6.8	598	T00509	probable vetispira
25	270.5	6.6	471	T06287	hypothetical prote
26	267	6.5	421	C96642	hypothetical prote
27	260	6.3	612	T05331	hypothetical prote
28	253	6.2	573	T05328	hypothetical prote
29	251	6.1	582	C71434	hypothetical prote

30	249	6.1	350	2	C56118	vetispiradiene syn
31	244.5	5.9	598	2	T05329	hypothetical prote
32	232.5	5.7	501	2	H70972	probable cyclase -
33	230.5	5.6	598	2	F96684	probable terpene s
34	228	5.5	946	2	T00024	ent-kaurene synthase
35	207.5	5.0	300	2	B56118	vetispiradiene syn
36	203	4.9	383	2	D71424	hypothetical prote
37	199.5	4.9	952	2	JC7227	hypothetical prote
38	157	3.8	2010	2	B71616	phosphatase (acid
39	156	3.8	1538	2	T29095	cardiac muscle fac
40	155	3.8	3351	2	T13812	lipophorin - fruit
41	151	3.7	753	2	S22802	transposase (clone
42	149.5	3.6	2335	2	T40186	probable phosphati
43	148	3.6	3394	2	T18501	hypothetical prote
44	147	3.6	2166	2	G70163	hypothetical prote
45	144	3.5	971	2	A70179	exodeoxyribonuclea

ALIGNMENTS

RESULT 1

T09672

ent-kaurene synthase B (EC 2.5.1.-) - winter squash

C:Species: Cucurbita maxima (winter squash)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004

C:Accession: T09672

R:Amaguchi, S.; Saito, T.; Abe, H.; Yamane, H.; Murofushi, N.; Kamiya, Y.

Plant J. 10, 203-213, 1996

A:Title: Molecular cloning and characterization of a cDNA encoding the gibberellin biosyn

A:Reference number: Z16814; MUID:96367664; PMID:8771778

A:Accession: T09672

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-789 <YAM>

A:Cross-references: UNIPROT:Q39548; EMBL:U43904; NID:G1431869; PIDN:AAB39482.1; PID:G143

A:Experimental source: immature seeds

C:Function:

A>Description: catalyzes the conversion of copalyl diphosphate to ent-kaurene

A:Pathway: gibberellin biosynthesis

A>Note: terpene cyclase

C:Superfamily: terpene synthase

C:Keywords: transferase

Query Match 49.2%; Score 2023.5; DB 2; Length 789;

Best Local Similarity 50.0%; Pred. No. 1.2e-112;

Matches 393; Conservative 150; Mismatches 212; Indels 31; Gaps 8;

Qy	16	NRPAALSAIHTASTSHGGQTNPTNLIIDT-----TKERIQKQFNVEISVSVD 64
Db	4	SRPTGVARFAASSSSSSSSASLFPQGVDDVTTTKTGAHFEEETKERIKKLFDKVELSVSAYD 63
Qy	65	TAWVAVPSPNSPKSPCFPECLNWLNNQNDGSLVNHNNHNNHLL-KDLSLSSTLAC 123
Db	64	TAWVAVPSPNSLNQPLFPFCINWVLSQADSGWLL-----HNDQLLMKANLLSTLAC 118
Qy	124	IVALKRWVVEDQINKGLSFIESNLASATKQSPGTGDIIPFGLLEYAKNLDINLLSK 183
Db	119	VLTKRWIGHDHMSKALDFIKSNIASATDENQSPVGFDIIPGMIEYAKNLDINLPLA 178
Qy	184	QTDPSLMHLKRELFQKCHSNEMDG--YLAVISGLNLYDMNNVKYQKNGSVFNSPS 241
Db	179	FTNVDAIVLRKRELELRSCRNSGGKAYLAYSEGIGKQDNDVMVMQYQRNGSLFNSPS 238
Qy	242	ATAAAFINHQNGPCINLYNSLDKFGNAVPTVYPHDLFIRLSMVDTTIRLGISHHVEI 301
Db	239	TTRAAFMHRNDGCFDYLRLQKFDGCVTTPDLYARLHMVDSLQKFGIARHFKHEI 298
Qy	302	KNVLDTYRCWVRDEQIFMDVVTCAAFALLRINGVEVSPDPLAETNBL-----AL 354
Db	299	RSVLDTYRCWQGEENIFLDASTCAMAFLMAYVEGYDVSSDQLTFQSEDIFFNCLGGYL 358
Qy	355	KDEVALETYHASHILYQEDLSGKQILK-SADFLKEIISTD---SNRLSKLHKVEENA 410

Db 359 KDFGASLELYKASQIITHPDSEVLNINSWTSRFLKHGLSSDSWSDRTDSVVKGEAYNA 418
 Qy 411 LKFPINTGLERINTRNRNIQLNVNDTRILKTTTHSSNISNTDYLRLAIVEDFTCOSIYRE 470
 Db 419 LEFFYNATLERLISKRAMESYSGDIVRISKSPYACINFGHQDFLELAIVEDFNTLQRIHLK 478
 Qy 471 ELKGLERWVVENKLDOLKPAROKTAYCYPVSAATLSSPELSDARISWAKNGILLTTVVDDF 530
 Db 479 ELEELQWVVENKLDLXKFFRLHGLGYCFAPAAATLTDPELHDARITAWAQNGVLTWVDDF 538
 Qy 531 PDIGTIDELTNLQICVKEWNVDDKCCSEHVRILFLALDKAICWIGDEAFKQOARDVT 590
 Db 539 YDGGSEBELDNLIELVEKWDPDGEGYCKDVEIVFLALHTVCEIGRRALVWGRSVM 598
 Qy 591 SHVIQTLWELMNSMLREAITWTRDAYVPTLNEYMENAYVSFALGPVIVKPAIYVFGPKLSEE 650
 Db 599 RNVIDGWLALLKMRKEAENSTKVVPSMEYMEQAHVSFALGPILPLFFVGPKLSEE 658
 Qy 651 IVSESEYHNLFLKMLSTOGLINDIHSFKREFEKGKLNALVALHLSNGESKVEEVEVEMM 710
 Db 659 MIGSEYQKLYKMLTAGLKNDIRSYDRECKEGLNLSLWMDGGGNVTKEEATEATK 718
 Qy 711 WMKKNRKLMLKLIPEENGSIYPRACKDAFWNMCHVLPNFYANDDGTGNTILDITVKDII 770
 Db 719 GDFERAIRLGLVLOEN-TTIPRACKDLFWKLSIVNLFYMEDDGTNSRLMNTVKAMF 777
 Qy 771 YNPLVL 776
 Db 778 EQPMDL 783

RESULT 2
 G96825
 hypothetical protein T8K14.12 [imported] - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
 C;Accession: G96825
 R;Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.P.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marziali,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A;Reference number: A86141; MUID:21016719; PMID:11130712
 A;Accession: G96825
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-785 <STO>
 A;Cross-references: UNIPROT:Q9SAK2; GB:AF005173; NID:94835764; PIDN:AAD30231.1; GSPDB:GN
 C;Genetics:
 A;Gene: T8K14.12
 A;Map position: 1
 C;Superfamily: terpene synthase

Query Match 49.1%; Score 2021; DB 2; Length 785;
 Best Local Similarity 52.6%; Pred. No. 1.7e-112;
 Matches 402; Conservative 130; Mismatches 208; Indels 24; Gaps 9;

Qy 34 QTNPTNLIIDTTERIKQKFNVEISVSDTAWAVPSPKSPFCECLNWLNNQ 93
 Db 27 QTRANNVSPFQTEKIRKMLEKVELSVAYSAYDTSNVAVPSPSQNAFLFPCCVKWLLDNQ 86
 Qy 94 LNDGSGVLNHTNNHHPKLDKLSSTLACTIVALKRWNVGDEQINKGLSFIESGLASATE 153
 Db 87 HEDGSGWLDNHDH---QSLKQVLSSTLALSKKGIQERQINKGLQFIELNSALVTD 143
 Qy 154 KSQSPFGFDIIFFGLELEYAK--NLDINLLSKQTFDSIMLHKRELBQKCHSNEM----D 207

144 ETIQKPTGFDIIFPGMIKYARDNLNITPLGSEVDD--MIRKRDLDLK-CDSEKFSKGR 200
 Qy 208 GYLAVISGLNLVDNMVKKYQKNGSVFNSPSATAAFINQHQPGLNLYNSLLDFG 267
 Db 201 AYLAVLGEGTNLKDWDLIIVKIQKNGSLFSPATTAATFQFGNDGCLVLCSLLOKFE 260
 Qy 268 NAVPTVYVPHDLFIRLSMVDTIERLIGISHHFRVEIKNVLDITYRCVWVERDEQIFMDVVTC 327
 Db 261 AAVPSVYFPDQVARIISIVTLBSLIGDRDKTEIKSILDETYRYWLRGDEBICLDLATCA 320
 Qy 328 LAFRLRLNGVEVSPDPLAETNELA-----LKDEYAALETTHASHILQEDLSGK 379
 Db 321 LAFRLLLAHGYDVSVDPLKPFABESGSDTLEGYKNTFSVLELKPAAQ-SYPHESALKK 379
 Qy 380 QILKASADFLKEITST--DSNRLSKLIHKEVENALKFPINTGLERINTRNRNIQLYN-VDNT 436
 Db 380 QCCWTKQYLEMELSSWVKTSVDRDKVLKEVEDALAFPSYASLERSDHRKILNGSAVENT 439
 Qy 437 RLTKTTHSSNISNTDYLRLAIVEDFTYTCOSIYREELKGLERWVVENKLDOLKPAROKTAY 496
 Db 440 RYTKTSYRLHNICTSDILKLAVDDEFNFCOSIHREMERLDRWIVENRLQELKFAKQCLAY 499
 Qy 497 CYFSAATLSSPELSDARISWAKNGILLTTVVDDFDDIGGTIDELTNLQICVKEWNVDDVK 556
 Db 500 CYFSAATLSSPELSDARISWAKNGILLTTVVDDFDDFVGGSKSELENLHLVEKWDLVGP 559
 Qy 557 DCCSEHVRILFLALDKAICWIGDEAFKQOARDVTSHVIQTLWELMNSMLREAITWTRDAYV 616
 Db 560 RYSSEHEVLIIFSRLDITLETGDKAFTYQGRNVTHHIVKILDLKLSMLREAEWSSDKST 619
 Qy 617 PTLINEYMENAYVSFALGPVIVKPAIYVFGPKLSEEVESSEYHNLFLKMLSTOGLINDIHS 676
 Db 620 PSLEDYMENAYISFALGPVIVPATYLGPPLEKTVDSHQYNQLYKLVNSTMGRLINDIQ 679
 Qy 677 FRREFEKGKLNALVALHLSNGESKVEEVEVEMMMKKNRKLMLKLIPEENGSIYPRAC 736
 Db 680 FKRESAEGKLNALVSLHMKHERDNRSKEVITIESKGLAERKRELHLKVLVEKGSVVPREC 739
 Qy 737 KDAFWNMCHVLPNFYANDDGTGNTILDITVKDIIYNPLVLNEN 780
 Db 740 KEAFLKMSKVLNLFYRKDDGFTSNDLMSLVKSVIYFVSLQES 783

RESULT 3
 T52059
 ent-kaurene synthase (EC 2.5.1.-) [imported] - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 09-Jul-2004
 C;Accession: T52059
 R;Yamaguchi, S.; Sun, T.; Kawaide, H.; Kamiya, Y.
 Plant Physiol. 116, 1271-1278, 1998
 A;Title: The GA2 locus of Arabidopsis thaliana encodes ent-kaurene synthase of gibberellin
 A;Reference number: Z25926; MUID:98205064; PMID:9536043
 A;Accession: T52059
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-785 <YAM>
 A;Cross-references: UNIPROT:O64952; EMBL:AF034774; PIDN:AAC39443.1
 C;Genetics:
 A;Gene: GA2
 C;Superfamily: terpene synthase
 C;Keywords: transferase

Query Match 49.1%; Score 2021; DB 2; Length 785;
 Best Local Similarity 52.6%; Pred. No. 1.7e-112;
 Matches 402; Conservative 130; Mismatches 208; Indels 24; Gaps 9;

Qy 34 QTNPTNLIIDTTERIKQKFNVEISVSDTAWAVPSPKSPFCECLNWLNNQ 93
 Db 27 QTRANNVSPFQTEKIRKMLEKVELSVAYSAYDTSNVAVPSPSQNAFLFPCCVKWLLDNQ 86
 Qy 94 LNDGSGVLNHTNNHHPKLDKLSSTLACTIVALKRWNVGDEQINKGLSFIESGLASATE 153

Db 87 HEDGSGLDNHDH---QSLKKOVLSTLASILAKKWKIGERQINQKGLQFIELNSALVTD 143
QY 154 KQSPSPIGFDIIIPFGLLEAYK--NLQINILSKQTDPSLMHLKRELBOKRCHSNEM-----D 207
Db 144 ETIQKTFGDIIPFGMIKYARDNLNIPLGSEVDD--MIRKQDLK-CDSEKFKSGRE 200
QY 208 GYLAYISEGLGNIYDNWVKYQKMGKSNVNSPATAAFAFINHONPGCLNLSLADKFG 267
Db 201 AYLAYILEGTRNLKQWDLIVKYQKNGSLFSDPATAAFTQPGNDGCLRLCSLLQKFE 260
QY 268 NAVPTVYHDLFLRLSWDTIELGLSHFRVEIKNVLDYETKVCVERBOIFMDVVTCA 327
Db 261 AAVPSVYPDQVARLSIIVLTLSEGLDRDFKTEIKSILDETYRVLWRGDEEICLDLATCA 320
QY 328 LAFRLRLNGYEVSPDPLAEITNELA-----LKDEYAALETYHSHILYQEDLSCK 379
Db 321 LAFRLLLAHGYDVSDPLKFAEESGFSDTLGYVKNVTSVLELFKAAQ-STPHESALKK 379
QY 380 QILKSAADFLKEIIST--DSNRLSKLIHKEVENALKFPINTGLERINTRRNIQLYN-VDNT 436
Db 380 QCCWTKQYLEMELSSWVKTSVRDKVLKKEVEDALAPPSVASLERSDHRKILNGSAVENT 439
QY 437 RILKTYHSSNTSDYLRSLAYEDEVTCOSIYREELKGLRWVVENKLDQLKFAKQKAY 496
Db 440 RYTKSYRLHNLCTSDIILKLVDDNFNFCOSIHREEMERLDWIVENRLQELKFAKQKAY 499
QY 497 CYFSVAATLSSPELSDARISWAKNGILTTVDVDFDGTIDELTNLQCVKKNVVDVK 556
Db 500 CYFSGAATLSPELSDARISWAKGVLTIVDDFVGVGSKELNLIHLVEKWDOLGVP 559
QY 557 DCCSHVRLFLKADACWIGDEAFKQOARDVTSHTVITQWLELNMNLMREAIWTRDAY 616
Db 560 EYSSEHVEIIFSVLRDITLFTGDKAFTYQGRNVTHIIVKWLDLKSLMRAEWSDDKST 619
QY 617 PTLNEMNAYVSPALGPVPAIYFVGPKLSEEVSEYHNLKLMSTQGRLLNDIHS 676
Db 620 PSLDYMNENAIISFALGPVIVPATYLIIGPEKTVSHQVQKLYKLVSTWGRLLNDIQQ 679
QY 677 FKRFKQKLNALVHLNGBSGKVEEVVEBEMMMIKNKKELMKLIFEENGSIIVPRAC 736
Db 680 FKRESAEKLNALVHLNGBSGKVEEVVEBEMMMIKNKKELMKLIFEENGSIIVPRAC 739
QY 737 KDAFWNMCHVNLFFVANDGFTGNTILDTVKDIIYNPLVNLNEN 780
Db 740 KEAFLKMSKVLNLFYRKDDGFTSNDLMSLVKSVIYEPVSLQES 783

RESULT 4
A96637
hypothetical protein Flp17.15 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: A96637
R;Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, X.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A96637
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-870 <STO>
A;Cross-references: UNIPROT:O22733; GB:AE005173; NID:G2443889; PIDN:AA871482.1; GSPDB:GN
C;Genetics:
A;Map position: 1
C;Superfamily: terpene synthase

Query Match 19.8%; Score 813.5; DB 2; Length 870;
Best Local Similarity 28.0%; Pred. No. 1.5e-40;
Matches 226; Conservative 163; Mismatches 322; Indels 97; Gaps 23;
QY 29 TSHGGGTNTNLIIDTTKERIQKQKNVE---ISVSSYDITAWVAVMPS-----PNSPKSP 80
Db 3 SYGSSNDLHAFVNEIKGEIQLSNINLPYFVSPSAYDTAWLSMIEEDINVDDELKP 62
QY 81 CPPECLNWLINNO-LNDGSGWGLVNHNNHLLKDS-----LSSTIACIALVK 128
Db 63 MFQGCGLDITMCNQNAREGF--MNST---SYTTVADGRDEDEKDMCILSTLACVALQ 117
QY 129 RNWVGEDQIKGLSPFESN---LASATEKSPSPIGFDIIPFGLLEYAKNLDINLSQ 184
Db 118 KWNIGCFHLHKTYRIERTNTEMIIGKYNIEGSPYRFAKFTGILELAQKGLHFPVS 177
QY 185 TDFSL---MLHRE---LEQKCHSNEMDGYLAYISEGLNLYDMN---MVKYQMKNG 234
Db 178 RCIEMLKGMFYQOEIIQREKLVHDCNYPKLLAYLEVLPKLYVTNQEDIIIVKSLDSMDG 237
QY 235 SVFNSPSTAAAFINHONPGCLNLSLDFKGNAPVTYP-HDLFIRLSWVDITIERLGI 293
Db 238 SLFQSPSATASAFMLTRNTKCLAYIQLVQKCPGVPQKYPINEDLIKLSWNLIESGL 297
QY 294 SHHFRVEIKNVLDYRCWVERD-----EQIFMDVVTICALAFRLRLNGYEVSPD 343
Db 298 GEFQIGIEHLVLEQVYSRYEKEKDFRMPMSYIADQLHKD---SLAFRLMRHGRDVSFR 353
QY 344 PLAEITNELALKDE-----YAALETYHSHILY--QEDLSSGKQ-----ILKSAFL 388
Db 354 SFCWFLNDQETRNHLERNIDSLVILSVYRATDLMFPGEHDLQAREYTRNLLEKRSI 413
QY 389 KEIISTDSNRLSKLIHKEVENALKFPINTGLERINTRRNIQLYNVNDTRILKTTVHSSNI 448
Db 414 KE--KWPPIYLNKIMHE-----LSTFWIARLKHLDHRMWIEDKSNVLSMEKASIRLHS 466
QY 449 SNTDYL-PLAVEDFTQOSIYREELKGLRWVVENKLDQLKFAKQKAYCYFSVAATLSS 507
Db 467 SYSOKLTHLAARNFQOAKYCRELEL-----TIEKTTYCYFATVTSLPY 512
QY 508 PELSDARISWAKNGILTTVDVDFDGTIDELTNLQCVKKNVVDVDCSEHVRILF 567
Db 513 EYALFKGLAKTALITIAIDFFDEKGSFNDLEGLTKAVLRWGBELK-----SYGNIIF 568
QY 568 LALKDAICWIGDEAFKQOARDVTSHTVITQWLELNMNLMREAIWTRDAYVPTLNEMNAY 627
Db 569 RALDDIVRETANTCRTHKTDIIVHLRNIGETTESWLRREAEWSKKGHTSSMDEYIRNGM 628
QY 628 VSPALGPVPAIYFVGPKLSEEVSEYHNLKLMSTQGRLLNDIHSFKREPKKGLIN 687
Db 629 ISIAAHATIALISICLMFPCFPHNKLKPGNVDYSITTLMLIIPRLNLDLQSYQKEQEQGKN 688
QY 688 AVALHLNGBSGKVEEVVEBEMMMIKNKKELMKLIFEENGSIIVPRACKDAFWNMCHVL 747
Db 689 SVLLHMKNHGLEIEDSIA-HIEKIDSKREFLEHVLNDGLSDLPKPKCEIHMSSCKV 747
QY 748 NFFVANDGFTGNT-ILDTVKDIIYNPL 774
Db 748 EMFENKKNRYDSNTEMLHDIKALYDPI 775

RESULT 5

T06783
ent-kaurene synthase A (EC 2.5.1.1) - garden pea
C;Species: Pisum sativum (garden pea)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C;Accession: T06783
R;Alt-Ali, T.; Swain, S.M.; Reid, J.B.; Sun, T.P.; Kamiya, Y.
Plant J. 11, 443-454, 1997
A;Title: The LS locus of pea encodes the gibberellin biosynthesis enzyme ent-kaurene synt
A;Reference number: Z15810; MUID:97260958; PMID:9107034
A;Accession: T06783

A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-801 <Alt>
A:Cross-references: UNIPROT:004408; EMBL:U63652; NID:G2160543; PIDN:AAB58822.1; PID:G2160543
A:Experimental source: cultivar Torsdag
C:Genetics: LS
A:Gene: LS
C:Function:
A:Description: converts the geranylgeranyl diphosphate into copalyl diphosphate
A:Pathway: the first step of the gibberellin biosynthesis
C:Superfamily: terpene synthase
C:Keywords: transferase

Query Match 17.6%; Score 725.5; DB 2; Length 801;
Best Local Similarity 27.0%; Pred. No. 2.3e-35;
Matches 224; Conservative 145; Mismatches 281; Indels 181; Gaps 32;

QY 14 KGNRPAALSAIHTASTSHGGQTNPTNL-----IIDTKER-----IQ 50
DB 39 KEKRCRAISKRTQEVGVQINVA TLKLSINVEDVIVIDBEEQDIRVGLVNIKISIL 98
QY 51 KQFKNVEISVSYDTAWAVPSPKSPCTPECLNWLINQNDGSLVNHVTHNH 110
DB 99 SSLEDEGITISAYDTAWVALVEDVAISTPQFPSSLEWIAKNQLOQSGWGDRLFSAH-- 156
QY 111 PLKDSLSSTLACIVALKWNVGCEQINKGLSFIEBNSLAS-ATEKSQSPGPIGDIIFPGI 169
DB 157 ----DRIINTLACVIALRSWNHSEKDKGMGPIFFRENLSKLENEHEHPIGFEVAFPSL 212
QY 170 LEYAKNLD-----INLASKQTFDFSLMLHRELEQKCHSNEMDGYLAISGLGN 219
DB 213 LEGARGIKLPCNDSPILKNIPEKDEKLTIRIPKEIMHK-----VPTLLHSLGMSG 266
QY 220 LYDNMVKYQKMGKSVFNSPSATAAFINQHPGCLNVLNLSLDKFGNAVPVYVPHDLF 279
DB 267 L-DWKQLLKQSQDGSFLSPSSTAFALMOTKGNCLKYNVNVKXFGNGVFNVPVDF 325
QY 280 IRLSMVDTIRLGLSHHFRVEIKNVLDITYRCWE-----RDEQIFMDVVTCALAFRL 333
DB 326 EHTVVDRLERLGLSRFRHEIKDCMNVYKIMSEKGI CWARNV- QDIDDTAMAFRL 384
QY 334 RINGVEVSPDLAEITNELAKDEY-----AALETYHSHLYQEDLSGKQI 381
DB 385 RLHGHQVS-----AHVFKHFERNGEFCFAGCQTAQVGMYNLFASQVLF-----PGEKI 435
QY 382 LK-----SADPLKEITSDNRL-----SKLHKVEVNAKFPINTGLERINTRNIQL 430
DB 436 LEHAKHFSAKVKE--KREANELDKWIIMKNLPEEVGYALDMPWYANLDRITRFYIDQ 493
QY 431 YNVD-NTRIILKTTVHSSNITDYLRLAVDFYTCQSIYREELKGLRWVVENKLDQLKF 489
DB 494 YGASDVWIGKTLRYMAYVNNNNLELAKDYNNCOAHLLENNVIGTWLSESLGEFGL 553
QY 490 ARQKATCYFSAVATLSSPELSDARISWAKNGILTVVDDFDIGGPIDELTLNIQCVEK 549
DB 554 SKRDLLEYLATGSIPEPERSHERLAWAK-----TTA-----LLETIKCYVR 596
QY 550 WNVVDVKDCC--SEHVRILFLAL-----KDAICWIGDEAFKQARDVTSH 592
DB 597 -NEDLRKDFAKFNDHDVDRDYSTARMKRNKTHSELVESLFATIGE--ISWDVRLSYGH 653
QY 593 VI-----QTWLELMNSMLRE--AIWTRDAVPTLNEYMENAVVSFALGPVVKPAIFYVG 644
DB 654 EIGYDMQCKWKLSSWSQSEGDCKEAGEALLIQLIN-LCSNHWIS-----EG 699
QY 645 PKLSEEVSEVHNLFKLMSTQORLNDI-HSPKRFKEGKUNAVAL--HLSNGRSGKV 701
DB 700 PSM-----QSTIQELL-----QLTNSICHKLSYQKQKELKGISQENITNSE---- 742
QY 702 BEEVVEENMMKRNKRLMKLIFEENGSIYPRACKDAFWMNCHVLNFFYA 752
DB 743 -----VESKMQELVQMVPQKCFNDIDFNVNKNTFFTIAK--SPFYA 780

RESULT 6

D85035
ent-kaurene synthetase A-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: D85035
R:anonymouse, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A>Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198
A:Accession: D85035
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-802 <STO>
A:Cross-references: UNIPROT:Q38802; GB:NC_001268; NID:G7269763; PIDN:CAB77763.1; GSPDB:G7269763.1
C:Genetics:
A:Gene: At4G02780
A:Map position: 4
C:Superfamily: terpene synthase

Query Match 17.2%; Score 709; DB 2; Length 802;
Best Local Similarity 32.6%; Pred. No. 2.2e-34;
Matches 174; Conservative 92; Mismatches 216; Indels 52; Gaps 13;

QY 30 SHGGQTNPTNLIIIDTTKERIQKQKNVEISVSYDTAWAVPSPKSPCTPECLNWL 89
DB 86 SVGSNSNAFKEAVKSVK-TILRNLTGDEITISAYDTAWVALIDA--GDKTPAPPSAVKWI 142
QY 90 INNQLNDGSLVNHVTHNHPLKDSLSSTLACIVALKWNVGCEQINKGLSFIEBNSLA 149
DB 143 AENQLSDGSGWDAYLFSYH-----DRLINTLACVVALRSWNLFPHQCKNGITFFRENIG 196
QY 150 S-ATERKSQSPGPIGDIIFPGLEAYAKNLDINLLSKQTFDFSLMLHRELEQKCHS---NE 205
DB 197 KLENDENDEHPGFEVAFPSLLEIARGINIDVPDPSVLKDIYAKKELKLTIRIPKEIMHK 256
QY 206 MDGYLAIYSEGKLVNWNVYKQKMGKSVFNSPSATAAFINQHPGCLNVLNLSLDK 265
DB 257 IPTTLHSLGVRDL-DWEKLLKQSQDGSFLSPSSTAFAPMOTRDSNLCLEYLRNAVKR 315
QY 266 FGNVATVYVPHDLFIRLSMVDTERLGLSHHFRVEIKNVLDITYRCWE-----RDEQI 319
DB 316 FNGGVFNVPVDFLFEHIWIVDRQLRGISRYFBEELKECLDYVHRITWDNGICWARCISHV 375
QY 320 FMDVVTCAAFRLRLRINGVEVSPDLAEITNELAKDEY-----AALETYHAS 367
DB 376 -QDIDDTAMAFRLRLRHQGVQVSADVFKNPEK-----GEFFCFVQSQNOAVTGMFNLYRAS 430
QY 368 HILYQEDLSGKQILKSADEPLKEIISTDSNRLSKLHK-----EVENALKFPINTG 418
DB 431 QLAPPRE-----BILKNAKEFSYNLLEKREBELDKWIIMKDLPGEIGFALEIPWYAS 485
QY 419 LERINTRNIQLYNVDN-TRILKTTVHSSNITDYLRLAVDFYTCQSIYREELKGLER 477
DB 486 LPRVETRFYIDQYGGENDVMWIGKTLRYMAYVNNNNLELAKDYNNCOAHLLENNVIGTWLSESLGEFGL 545
QY 478 WVENKLDQLKFAKQKATCYFSAVATLSSPELSDARISWAKNGILTVVDDFF 531
DB 546 WYENRLSEWGVRSSELLECYLAATIPFSERSHERMVMWAKSVLVKAISSSF 599

RESULT 7
T02959
kaurene synthase A - maize
C:Species: Zea mays (maize)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004
C:Accession: T02959
R:Bensen, R.J.; Johal, G.S.; Crane, V.C.; Tossberg, J.T.; Schnable, P.S.; Meeley, R.B.;
Plant Cell 7, 75-84, 1995
A>Title: Cloning and characterization of the maize An1 gene.
A:Reference number: Z14794; MUID:95210929; PMID:7696880
A:Accession: T02959

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-823 <BEN>

A;Cross-references: UNIPROT:Q41771; EMBL:L37750; NID:g576885; PIDN:AAA73960.1; PID:g5768

C;Genetics:

A;Gene: AN1

C;Function:

A;Description: cyclase involved in gibberellic acid biosynthesis

C;Superfamily: terpene synthase

Query Match 16.0%; Score 657.5; DB 2; Length 823;
Best Local Similarity 30.5%; Pred. No. 2.7e-31;
Matches 192; Conservative 106; Mismatches 266; Indels 65; Gaps 22;

```
QY 12 LTKSRPAAIAIHTASTSHGGTNPNTLIIDTKERIOQKFNKVEISVSSYDTAWAMV 71
Db 70 LNRGRKPDLEBHQABEA---ELQP---LIQVRAML-RSMNDGUTSASAYDTAWAMV 122
QY 72 P--SPNSPKSPCFPECLNMLINNQLNDGSGVLNHTNHNHPLK--DSLSTLACIVAL 127
Db 123 PKVGGDGAQPOFPATVRMIVDQLPDGSGW-----DSALFSAYDRMINLACVVAL 174
QY 128 KMWVGEQIQKGLGFIENL-ASATEKSOPSPIGFDIIFPGLELLEYAKNLDINLLSKQTD 186
Db 175 TKWLEPARCEAGLSFLHNMMWLAEAESEMPIGFEIAPPSLIQTARDLGV-----VD 228
QY 187 F-----SLMLHKLELEQKRCHSNEMDGY---LAVISEGLNLYDNNMKVQMKNGSV 236
Db 229 FPGHPALQSIYANREVKLKIPRDMHVRVPSIIHSEGMPLD-DWPRLLMLQSCDGSF 287
QY 237 FNSPSATAAFINHQNPGLNVLNGLDKFGNAVPTVYPHDIFIRLSMVDITIERLGISHH 296
Db 288 LFSPSATAYALMOTGDKCKFEYIDRVKFGNGGVENVYVDLFEHVVVDRLERLIGISRY 347
QY 297 FRVEIKNVLDEYR-----CWVERDQIFMDVVTICALAFRLIRINGEYVSPDLAEIT 349
Db 348 FQREIEQCNDVYNRHWTEDGICWARKSN--VKVDVDTAMAFELLLHGVNYSVPSVFNKE 405
QY 350 NE-----LALKDEVAALAEYH---ASHILYO-EDLSSGKQILKSADFLKEI-----ISTDS 396
Db 406 KDGREFFCVGQSTQAVTGMYNLRASQISFQEDVLHRAVP-SYEFRLQREQGMIRDK 464
QY 397 NRLSKLIHKVENALKFPINTGLERINTRNIQLY-NVDNTRILKTTYHSSNISNTYLR 455
Db 465 WIVAKDLFGEVQYTLDFPWAYSLPRVEARTYLDQYGGKDDVWIGKTLXRMPLVNDTYLE 524
QY 456 LAVDEFTYCQSIYREELGLERWVENKLDQKFAKQKTAICYFSVAATLSPPELSDARI 515
Db 525 LAIRDFFNECOALHQEGLQTYWKNGLDAPGVEPDVLRSYFLAAACAFEPSSRAERL 584
QY 516 SWAKNGILTTVVDDFF-DIGGTIDELTNLIQCVEKKNVDVDKCCGSEHVRILFLALKDAI 574
Db 585 AWARTSMIANAISTHLRDISDKKLECFVHCLYEEN-DVSWLKNPNPDVILERALRLKI 643
QY 575 CWIGDEAFKQW--ARDVTSHVQIOWLELM 601
Db 644 NLLAQEALPIHEGORFIHSLSLAWTEWM 672
```

RESULT 8

H84633

probable limonene cyclase [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004

C;Accession: H84633

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon,
L.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: H84633

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-591 <STO>

A;Cross-references: UNIPROT:Q9ZUH4; GB:AE002093; NID:g4115381; PIDN:AA03382.1; GSPDB:GN

C;Genetics:

A;Gene: At2g24210

A;Map position: 2

C;Superfamily: vetispiradiene synthase 1

Query Match 10.9%; Score 446.5; DB 2; Length 591;
Best Local Similarity 27.0%; Pred. No. 5.9e-19;
Matches 138; Conservative 97; Mismatches 242; Indels 35; Gaps 14;

```
QY 281 RLSDMTTIERLIGSHHVEIKNVLDETYRCWVERDEQIFMDVVTICALAFRLIRINGEYV 340
Db 98 QLEFIDDDLOKGVSHFEAEIDNILTSSYK--KDRNTIQESDLHATALEFRFLRQHGPNV 155
QY 341 SP---DPLAEITNELALKDEYAALETYHSHILYQEDLSSGKQIL---KSADFLKEIIST 394
Db 156 SEDVDFVFMENCGKFPDRDDIYGLISLYEASYL--STKLDKNLQIFIRPPATQOOLRDFVDT 213
QY 395 DSNR--LSKLIHKVENALKFPINTGLERINTRNIQLYNVDNTRILKTTYHSSNISNTD 452
Db 214 HSNEDFGSCDMVEIVVQALDMFYWMRLSTRWIDVYG-----KRONYKNLV 262
QY 453 YLRLLAVEDFTYCQSIYREELGLERWVENKLDQKFAKQKTAICYFSVAATLSPPELS 511
Db 263 VVEFAKIDFNIVQALHQEELKNVSSMMWETGLKOLYFARDRIVENYFTWIGIQIOBPQYG 322
QY 512 DARISWAKNGILTTVVDDFFDIDGTIDELTNLIQCVEKKNVDVDKCCGSEHVRILFLALK 571
Db 323 YVRQMTKINALTTIDDIYDGLLEELQLFVAFENWDIN-RLELPEYNELCFVLIVY 381
QY 572 DAICWIGDEAFKQWQARDVTSHVQIOWLELMNSMLREAIWTRDAYVPTLNEMENAYVSPA 631
Db 382 NEVNSIACEILRTKNINVIPFLKKSNTDVSXAVLVEAKWKSCHKPNLEEYMQNARISIS 441
QY 632 LGPIVKPAIYFV-GPKLSEETVESSEYH--NLFKLMSTQGRLLNDIHFKEFEKGLK-N 687
Db 442 -SPITFVHFYCVFSDQLSIOVLETLSSHQQNVNRCSSVFRFLANDLVTSPPELARGDYCK 500
QY 688 AVALLHLSGSGKVEEVEEVEEMMMKMKRKLKELKLIPEENGSIIVPRACKAFNMCHVL 747
Db 501 SIQCVMSS--ETGASEDKARSHVRQWINDLWDMENYKWAHSSILHHDPMETVINIARMS 558
QY 748 NFFVANDDGFTG---NTILDFTVKDIIYNPLVL 776
Db 559 QCMYQYGDGSGSPEKAKIVDRVMSILFNPPIEL 590
```

RESULT 9

F71434

probable limonene cyclase - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

A;Variety: columbia

C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 09-Jul-2004

C;Accession: F71434

R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirks
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giele
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998

A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech,
erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ansc
C.; Chaiwalaks, N.

A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thali

A;Reference number: A71400; MUID:98121113; PMID:9461215

A;Accession: F71434

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-600 <BEV>

A;Cross-references: UNIPROT:O23516; GB:Z97341; NID:g2244991; PID:g2245028

C;Genetics:

A;Map position: 4COP9-4G3845

C;Superfamily: vetispiradiene synthase 1

Query Match 10.5%; Score 431; DB 2; Length 600;
 Best Local Similarity 24.6%; Pred. No. 5e-18;
 Matches 154; Conservative 116; Mismatches 279; Indels 78; Gaps 17;

180 LLSKOTDPSLMI---HKRELEOKRCHSNEMD---GYLAYISEGLNLYDMNWKYQMK 232
 22 LLSNNTLALTELELCHLSLSDAKHCHORNERNORRHGESPSQERLISLIDKNVRLITM 81
 233 GSVNFSPSATAAFINHQPCGLNLSLLDFGNVAVTVYPHDLF-IRLSNMVDTIERL 291
 82 SGG-----GVLYKDCGRKEVETAEKAI---LFXELEMIDSLQRL 119
 292 GISHFRVRIKVLDETY--RCWVERDEQIFMDVVTTCALAFRLIRINGVEVSPDLAET 349
 120 GISVYHKEIHDLRIKHQHGEIERETO---DLHATSLEFILLRQHGFDFVQADPVTI 176
 350 NELA-----LKDEYAALFTYHSHILYQEDLSSGKQILKSAFLKEIISTDSNLSK-- 401
 177 SETGFERTLHSDIKGLLSLYEASYSMDSEFKLRETRIVANKRULSEFVAESSKTICRED 236
 402 --LIHKEVENALKFPINTGLERINTRNRNIQLYNVDNTRILKITYTHSSNISNTDYLRLAVE 459
 237 ETVILLEWVKRALETPYHWSIRRLERAWYINV-----YEKKHEMNPLLEFAAI 284
 460 DFTQCSIVREBLK-----GLERWVENKLDQKFARQKTAICYFSVAATLSSPELSDAR 514
 285 DFNMLQANQHEELKLSLSTGL-----MKQLDFVRDRITESYFWTIGIFYEPEFYCYR 336
 515 ISWAKNGILTVVDDFDIGTIDELTNLIQCVEKNVND-VDKDCCSEHVRILFLALKDA 573
 337 KILTKLFMLIVMDIYDYGTELELELFTNVVEKDVNHER--LPNYMEMCFLYNE 394
 574 ICWIGDEAFKQARDVTSVHIQVWLELMSMLREAIWTRDAYVPTLNEYMENAVYFALG 633
 395 INQIGYDVLKDLGNLIPYKQVWTDLFTLFTESKWYTKHGPSPEEYMQNGVIVSSVP 454
 634 PIVKPAIYVGPKLSEI-IVESSEYHNLFKLMSTQGLLNDIHSFKREFKQKLNVALH 692
 455 TILLHLFSLSDHISDQTLTDDSKNHSVVRSCATILRLANDLATSTEEMARGD-SPKSQV 513
 693 LSNESGKVEEVEEVEEMMMKMKELMKLIFEENGSIIVPRACKDAFNMCHVNLFFYA 752
 514 CYMYETRASEEEARRHMQSMISDWDIINSDLKTAHTSLSIPRGFLAANLNRVQCIYR 573
 753 NDGFTG---NTILDVTKDIYINPLVL 776
 574 HGDGHSPEKTKVDYIQSVLFPVPL 600

RESULT 10
 T08174
 sesquiterpene cyclase (BC 2.5.1.-) - pepper
 C;Species: Capsicum annuum (pepper)
 C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
 C;Accession: T08174
 R;Back, K.; Shin, D.H.; He, S.
 Plant Cell Physiol. 39, 899-904, 1998
 A;Title: Cloning and bacterial expression of sesquiterpene cyclase, a key branch point e
 A;Reference number: Z16395; MUID:99033462; PMID:9816674
 A;Accession: T08174
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-559 <BAC>
 A;Cross-references: UNIPROT:O65323; EMBL:AF061285; NID:G3108342; PIDN:AAC61260.1; PID:G3
 C;Function:
 A;Pathway: the synthesis of phytoalexin capsidiol
 A;Note: unduced by UV
 C;Superfamily: vetispiradiene synthase 1
 C;Keywords: isoprenoid biosynthesis; transferase

Query Match 10.4%; Score 426; DB 2; Length 559;
 Best Local Similarity 25.3%; Pred. No. 9e-18;

Matches 129; Conservative 102; Mismatches 204; Indels 74; Gaps 15;

282 LSNVDTIERLGISHFRVEIKNVLDITYRCWVERDEQIF-MDVVTTCALAFRLIRINGVEV 340
 73 LMLIDVIERGLIAYHFEKEIDEILDRITNENSPFEDVYNEDELCTCRQLRQLRHGYNI 132
 341 SPDPILAETITN-----ELALKDEYAALFTYHSHILYQEDLSSGKQILKSA-DFLKEIIS 393
 133 SLKIFSKFDGNGRLKESLASDVLGLLSLYEASHV-----RSHGEDILDELALAFSTHLE 187
 394 TDSNRLSKLIHKEVENALKFPINTGLERINTRNRNIQLYNVDNTRILKITYTHSSNISNTDY 453
 188 SATPHLEYPLKEQVRHALESQSLHKGIPRI-----EIQFF-----ISSYVDKQAIKNVDL 236
 454 LRLAVEDEFTQCSIVREBLKGLERWVE-NKLDQKFARQKTAICYFSVAATLSSPELSD 512
 237 LRFALKDYNMLQMLHKQELAEVSRWVKDLNFVNTLYPARDRVVVCYFWALGVYVEPQYSQ 296
 513 ARISWAKNGILTVVDDFDIGTIDELTNLIQCVEKNVNDKCCSHVRLFLALKD 572
 297 ARVMLVKTIAMTSIVDDTYDAYGTVDLAIYTDVIQWMDIK-BIDSLPDYMKISYKALLD 355
 573 AICWIGDEAFKQARDVTSVHI---QVWLELMSMLREAIWTRDAYVPTLNEYMENAVY 628
 356 ----LYKDYKEKMSRDSRSHVVYAKERLKLIVKSYNIEAKWFIENGHPPASLEYLRNAV 411
 629 SPALGPIVPAIYVGPKLSEI-----IVESSEYHNLFKLMSTQGLR 670
 412 TTTYVYIANTS--YLGMKYAKSQQEPWLSKNPKILGCVTICRVIDDIAITYEVEKNRGQL 469
 671 LNDIHSFKREFKQKLNVALHLSNGESGKVEEVEEVEEMMMKMKELMKLIFEENG 730
 470 STGIECYMDYSVSTKEAMAKFOEMGESG--WKDINEGML-----RPT 510
 731 IYPRACKDAFNMCHVNLFFYA-NDGFT 758
 511 PIPMEFLSRILNARLVDVTVKHNEGYT 539

RESULT 11
 A48863
 limonene cyclase - spearmint
 C;Species: Mentha spicata (spearmint)
 C;Date: 12-May-1995 #sequence_revision 19-May-1995 #text_change 09-Jul-2004
 C;Accession: A48863
 R;Colby, S.M.; Alonso, W.R.; Katahira, E.J.; McGarvey, D.J.; Croteau, R.
 J. Biol. Chem. 268, 23016-23024, 1993
 A;Title: 4S-limonene synthase from the oil glands of spearmint (Mentha spicata). cDNA
 A;Reference number: A48863; MUID:94043077; PMID:8226816
 A;Accession: A48863
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-599 <COL>
 A;Cross-references: UNIPROT:Q0322; GB:L13459; NID:G410229; PIDN:AAC37366.1; PID:G410230
 C;Superfamily: vetispiradiene synthase 1

Query Match 10.0%; Score 413; DB 2; Length 599;
 Best Local Similarity 24.5%; Pred. No. 5.9e-17;
 Matches 135; Conservative 115; Mismatches 209; Indels 92; Gaps 19;

256 LNVNLSLLDKFGNAVTVYPHDLFI-----PLSMVDITIERLGISHFRVEI 301
 70 VNFQTSLSDYKEDKHVIRASELVTLVKMELEKETDQIRQLLEIDDLQWGLSDHFOFNEF 129
 302 KNVLDETYRCWVERDEQIF-----MDVVTTCALAFRLIRINGVEVSPDLAETITNE-- 351
 130 KEILSSIY-----LDHHYKNPFPKEERDIYSTSLAFRLIREHGFQVAQEVDFSKNEEG 184
 352 ---LALKDE-YAALFTYHSHILYQEDLSSGKQILKSA-----DFLKEIIS---TDSNRL 399
 185 EFKESLSDTRGLLQLLYEASFL-----LTEGETTLESAREFATKFLKEKNEGGVDGLL 239
 400 SKLIHKEVENALKFPINTGLERINTRNRNIQLYNVDNTRILKITYTHSSNISNTDYLRLAVE 459

```
Db 240 TRIAY-----SDDIPLHWRIK-----FNAPVWIEWYKRPDMNFWLELAL 282
QY 460 DFTYQTSIVREELKGLERWVVENK-LDQLKEARQKTAYCYPFVAATLSPFLSDARISWA 518
Db 283 DLNITVQAQOEELKESFRWRNTGVEKUPFFARDRLVECYFWNTGTIIEPRQHASARIMMG 342
QY 519 KNGILTTVVDDFFDGGTIDELTNLIQCVEKWNVDVDDCCSEHVRILFLALKDAICWIG 578
Db 343 KVALITVIDDIYDVVGTLELEEQFTDLIRWDIN-SIDQLPDYQWQCFALANNFVDDTS 401
QY 579 DEAFQWQARDVTSVHIQWLELMNMLREAIWTRDAYVPTLNEMENAYVSFALGPIVXP 638
Db 402 YDVMEKGVNVPYLRQSWVDLADKYMVEARFYGCHKPSEYLENSWQIS-GFCMLT 460
QY 639 AIYF-VGPKLSEIIVES-SEYHNLFKLMSTQGRLLNDIHSFKREFKGL-NAVALHLSN 695
Db 461 HIFRVTSFKEYETVDSLYKYHDLVRWSFVURLADDIGTSVEEVRGDPVKSLOCYMSD 520
QY 696 GESGKVEEVEEEMMMIKNRKELMKLIFEENGSI-VPRACKDAFW-----NMCH 745
Db 521 YNASEAE-----ARKHVWLLIAEWWKMAERYSKDSPPFGKDFIGCAVDLGR 567
QY 746 VLNFPYANDDG 756
Db 568 MAQLMYHNGDG 578

RESULT 12
G71434
Probable limonene cyclase - Arabidopsis thaliana
A:Variety: Columbia
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 09-Jul-2004
C:Accession: G71434
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirks
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
C.; Chalwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A:Reference number: A71400; MUID:98121113; PMID:9461215
A:Accession: G71434
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1024 <BEV>
A:Cross-references: UNIPROT:O23517; GB:Z97341; NID:g2244991; PID:e327011; PID:g2245029
C:Genetics:
A:Map position: 4COP9-4G3845

Query Match 9.9%; Score 406; DB 2; Length 1024;
Best Local Similarity 25.3%; Pred. No. 3.5e-16;
Matches 141; Conservative 104; Mismatches 226; Indels 86; Gaps 18;

QY 278 LFIRLSMVDTRIERIGISHHFRVEIKNVLDFTY-----RCWVER-DEQIFMDVVTTCALAFRL 332
Db 27 LLEQLELIDTLQRLGVSHFEQEIKKTITNVHVKNVRAHKNRINDRWGLIYATALEFRL 86
QY 333 LRINGVSVSDPLAETINELALKDEYALETYHSHILYQEDLSSGKOILKSADFLKEII 392
Db 87 LR--QHDVFD---GNIGVLDLDDKIKGILSLIYEASYLSTRIDTKLESIVYTKRLAKFV 141
QY 393 STDSNRL-SKLHKEVENALKFPINTGLERINFRNTQLYVNDTRILKTYTHSSNLSNT 451
Db 142 EVKNETKSYTLRKNWVTHALEMPYHRVGRLEARWYIEVYG-----ERHDMNP 189
QY 452 DYLRAVEDFTYTCOSIYREELKGLERWVVENKLDQ-LKEARQKTAYCYFVSVAATLSSPEL 510
Db 190 ILLELAKLDFNFVQAIHQDEKLSLSSWSKGLTKHLDFDRDRTGTEGYFSSVGVWVPEF 249
QY 511 SDARISWAKNGILTTVVDDFFDGGTIDELTNLIQCVEKWNVDVDDCCSEHVRILFLAL 570
```

```
Db 250 AYHRQWMLTKVFMILITIIDDIYDYGLEELQFTTIVEKWNVRLEE-LPNTMKLQCLFL 308
QY 571 KDAICWIGDEAFKQWQARDVTSVHIQWLELMNMLREAIWTRDAYVPTLNEMENAYVSF 630
Db 309 VNEINQIGVFVLRDKGFNVIPVYKESWADMCITFLKEAKWYKSGYKFPNFEYQWQWISS 368
QY 631 ALGPVKAIPAVPGKLSSEIVE--SSEYHNLFKLMSTQGRLLNDI-----HSP----K 678
Db 369 SV-PTILHLFLCL--LSDQTLIDILGYSNHSVSRSSATILRLANDLATSSVSHGFTTYNT 424
QY 679 REFKEG-KLNAVALHLSNGESGKVEEVEEEMMMIKNRKELMKL1-----FEENG 729
Db 425 BELARGDTMKSVOCHWH--FTGASEAE-----SRAYIQIIGLVAMDLDLNNKKS 471
QY 730 SIVPRACKDAFWNMCHVLNFFYANDDG-----FTGNTILDTV 766
Db 472 CRLHQGFLEAAANLGRVAQCVQYGDGHCPCDKAKTVNHNRRQRIKWRNAFLGSGIMTTI 531
QY 767 KDIIYNPLVLYNENEEO 783
Db 532 VVLLHTPTFSVFSDEEE 548

RESULT 13
T03714
5-epi-aristolochene synthase - common tobacco
N:Alternate names: sesquiterpene cyclase
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004
C:Accession: T03714
R:Facchini, P.J.; Chappell, J.
Proc. Natl. Acad. Sci. U.S.A. 89, 11088-11092, 1992
A:Title: Gene family for an elicitor-induced sesquiterpene cyclase in tobacco.
A:Reference number: Z15024; MUID:93066390; PMID:1438319
A:Accession: T03714
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-550 <FAC>
A:Cross-references: UNIPROT:Q40577; EMBL:L04680; NID:gi70342; PID:AAA19216.1; PID:g50551
A:Experimental source: strain NK326
C:Genetics:
A:Introns: 37/3; 126/1; 251/2; 324/2; 370/3; 452/3
C:Function:
A:Description: mediates the conversion of the isoprenoid intermediate farnesyl diphosphat
C:Superfamily: vetispiradiene synthase 1

Query Match 9.7%; Score 399; DB 2; Length 550;
Best Local Similarity 23.7%; Pred. No. 3.6e-16;
Matches 123; Conservative 109; Mismatches 216; Indels 70; Gaps 16;

QY 282 LSWVDTRIERIGISHHFRVEIKNVLDFTYRCWVERDEQIFMDVVTTCALAFRLRINGVEVS 341
Db 70 LNLIDIIERIGISVHFEKEIDEILDQIY-----NQNSCNCDLCTSAQLRLLRQHGNI 124
QY 342 PDPLAETNE-----LALKDEYALETYHSHILYQEDLSSGKOILKSADFLKEI-IS 393
Db 125 PEIIFSKPDENGKFKESLA-SDVLGLNLNLYEASHVRTHAD-----DILEDALAFSTHLE 178
QY 394 TDSNRLSKLHKEVENALKFPINTGLERINFRNTQLYVNDTRILKTYTHSSNLSNTDY 453
Db 179 SAAPHLKSPREQVTHALEQCLHKGVPVETR-----FFISSIYDKQSQKNVL 227
QY 454 LRLAVEDFTYTCOSIYREELKGLERWVVE-NKLDQLKPARQKTAYCYFVSVAATLSSPLSD 512
Db 228 LRPAKLDFNLLQMLHQELAQVSRWKKLDLDFVITLTPARDRVVECVFWALGVFEQYSQ 287
QY 513 ARISWAKNGILTTVVDDFFDGGTIDELTNLIQCVEKWNVDVDDCCSEHVRILFLALKD 572
Db 288 ARVMLVKTISMSISIVDDTFDAYGTVKELEYATDAIQWDLIN-EIDRLPDVWKISYKAILD 346
QY 573 AICWIGDEAFKQWQARDVT---SHVITQWLELMNMLR-----EAIWTRDAYVPTLNEME 624
```

```
Db 347 -----LYKDYKELSSAGSHIVCHAIERMKVEVRYNVVSTWTFIEGYMPPVSEYLS 398
Qy 625 NAYVSFALGPVIVKPAIYFVGPKLSEIEVSSYHNLFLKLMSTQGRLLNDIHSFKRFBKGG 684
Db 399 NALATTTYYLATTGYLVGKMSATEQDFEWSKNPKILEASVIIICRVDDTATYVEKSRG 458
Qy 685 KLNALVALHLSNGESKVEEVEVMMMLKKNRKLMLKLIPEENGSIIVPRACKDAF--- 740
Db 459 QT-ATGIECCMRDYGISTKEMAKQFQMAETAWKDI-----NEGLLRPTPVSTETLPTI 511
Qy 741 WNMCHVLNFFYA-NDGFTG-----NTILDTVK 767
Db 512 LMLARIVEVYTHNLDGTHPEKVLKPHIINLLVDSIK 549

RESULT 14
vetispiradiene synthase 1 - Hyoscyamus muticus (fragment)
C:Species: Hyoscyamus muticus
C>Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 09-Jul-2004
C:Accession: A56118
R:Back, K.; Chappell, J.
J. Biol. Chem. 270, 7375-7381, 1995
A:Title: Cloning and bacterial expression of a sesquiterpene cyclase from Hyoscyamus muticus
A:Reference number: A56118; MUID:95221394; PMID:7706281
A:Accession: A56118
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-520 <BAC>
A:Cross-references: UNIPROT:Q39978; GB:U20188; NID:G763422; PIDN:AAA86337.1; PID:G763423
C:Superfamily: vetispiradiene synthase 1

Query Match 9.7%; Score 398.5; DB 2; Length 520;
Best Local Similarity 23.9%; Pred. No. 3.5e-16;
Matches 127; Conservative 115; Mismatches 224; Indels 65; Gaps 15;

Qy 260 NSLLDKFGNAVPIVYPH-----DLFRLSMVDTERIGISHHFRVEIKNVLDET 308
Db 3 NOVAEKYAEIETLKQGTSTMLSAAGTTTETKINLIDIERGLIAHYHEKQIEDMLDHI 62
Qy 309 YRCWVERDEQIFMDVVTICALAFRLRLRINGVEYSPDFLAETN-----ELAKDEYAALE 362
Db 63 YRADPYFEAHEYNDLNTSSVQFELLRQHGYNVSENFPSRFQDANGKFKESLRSDIRGLLN 122
Qy 363 TYHASHI-LYQEDLSSGKQLKADFLKEIISDLSNRSLKLIHKEVENALKFPINTGLER 421
Db 123 LYEASHVTRHKEDI-----LEEALVFSVGHLESAAPHLKSPLSKQVTHALEQSLHKSIPR 177
Qy 422 INTRMIQLYNDVTRILKTYTHSSNITSNTDYLRLAVEDFYTCQSYREELKGLERWVVE 481
Db 178 VEIRYFISI-----YEEEPKNDLLRLPAKLDYNLIQLHLKHELSEVSRWKD 225
Qy 482 -NKLDQLKFARQKTAYCYFSAATLSPELSDARISWAKNGIILTVVDDPFDIGGTIDEL 540
Db 226 LDFVTLLPYARDRAVECYFTWGVYAEQYSQARVMLAKTIAMISIVDDTFDAYGIVKEL 285
Qy 541 TNLIOQVEKNVDVDDKCCSEHVRILFLAKDALCIGDEAFKQWQADVTSHVIQIOWLEL 600
Db 286 EVYTDATQRWDIS-QIDRLPEYMKISYKALLDYDDYEKELSKDGRSDVHYAKERKKEI 344
Qy 601 MNSMLREAIWTRDAYVPTLNEYMENAVVSA-----LGPVVKPAIYFVGPKLSEI 651
Db 345 VGNFYIEGKFIKGYMPSVESEYLSNALATSTYYLITTSYLGKMSATKEHEFWLATNPKI 404
Qy 652 VESSEYHNLFKLMS TOGRLLNDIHSFKRFBKGL-NAVALHLSN-GESEKVEEVEVVEE 709
Db 405 LEAN-----ATLCRVDDIATYVEKRGQIATGIECYMRDYGVSTEVAMEKPFQEM 455
Qy 710 M-MMKKNRKLMLKLIPEENGSIIVPRACKDAFNMCHVLNFFYA-NDGFT 758
Db 456 ADIAWKDVNEILRPT-FVSSEILTR-----ILNLARIIDVYTKHQDGYT 500
```

```
RESULT 15
S68366
(-)-delta-cadinene synthase isozyme XC14 - Gossypium arboreum
C:Species: Gossypium arboreum
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C:Accession: S68366
R:Chen, X.Y.; Chen, Y.; Feinstein, P.; Davisson, V.J.
Arch. Biochem. Biophys. 324, 255-266, 1995
A:Title: Cloning, expression, and characterization of (+)-delta-cadinene synthase: a cat-
A:Reference number: S68365; MUID:96132653; PMID:8554317
A:Accession: S68366
A:Molecule type: mRNA
A:Residues: 1-554 <CHE>
A:Cross-references: UNIPROT:Q39760; EMBL:U23205; NID:G1045313; PIDN:AAA93065.1; PID:G1045
A:Experimental source: cultivar Nanking
C:Superfamily: vetispiradiene synthase 1
C:Keywords: phytoalexin biosynthesis

Query Match 9.1%; Score 373; DB 2; Length 554;
Best Local Similarity 24.6%; Pred. No. 1.3e-14;
Matches 131; Conservative 98; Mismatches 217; Indels 86; Gaps 20;

Qy 281 RLSDMTTERIGISHHFRVEIKNVLDFTYRCWVERDEQIFMDVVTICALAFRLRLRINGYEV 340
Db 73 KLAFTDSVQRLGVSYHFTKEIDELENIY-----HNNDAENLDYTTSLRFLRREHGFNV 128
Qy 341 SPDLAEITNE-----LALKDEYAALETYHSHILYQEDLSSGKQLKLSADFLKEIIST 394
Db 129 SDGVNFKDEQGNKSSVTSVDRGLLELYQASLYR-----VHGEDIIDEAISF 177
Qy 395 DSNRLSKLI-----HKEVENALKFPINTGLERINTRRNIIQLYNDVTRILKTYTHSSN 447
Db 178 TTNHLASVASLDYPLSEVSHALKQSIRRGLPVREARHYL-----SVYQDIE 225
Qy 448 ISNTDYLRLAVEDFYTCQSYREELKGLERWVVENKLD--OLKFAEQKTAYCYFSAAT 504
Db 226 SHNKVLLEFAKIDFNVMQVLLHRRKRLSEISRWKD--LDFQRKLPYARDRVVEGYFWISGV 283
Qy 505 LSSPELSDARISWAKNGIILTVVDDPFDIGGTIDELTNLIQCVERKNVD-VDXDCCSEHV 563
Db 284 YFEPQYSLGRKMLTKVIAMASIVDDYDSYATVEELIPTYKALERWDIKIDE--LPEYM 341
Qy 564 RILFLALKDAICWIGDEAFKQWQADVTSHVIQ-----TWLELMNSMLREAIWTRDAYVPTL 619
Db 342 KPSYKALLD---VYEEMEQLVAKHGRQYREYAKNAMIRLAQSYLVEARWTLQNYKPSF 397
Qy 620 NEYMENA-----YVSA-----LGPVVKPAIYFVGPKLSEIEVESSEVENLFKLMSTQ 668
Db 398 EEFKANALPTCGYAMLAITSFVGMGDIVTPET-FKWAANDPKLIQAS-----TIIC 447
Qy 669 RLILNDI--HSFKRFBKGLNAVALHLSNGESKVEEVEVVEEVMNMIMKKNRKLMLKLIPE 726
Db 448 RFMDDVAEHKFKER-REDDCSALECYME--EYGTAGEAYDVFNKGVESAWKDVNKEFLK 504
Qy 727 ENGSIVPRACKDAFNMCHVLNFFYANDDGFT--GNWILDTVDKDIYNPLVL 776
Db 505 P--TEMPTVELNRSNLARVMDVLYREGDGYTVVGKAAKGITSLTLEPVAL 554
```

Search completed: October 17, 2004, 08:54:21

Job time : 53.1457 secs

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	2023.5	49.2	789	3	US-08-727-308-1	Sequence 1, Appli
2	1523	37.0	732	4	US-09-614-912-24	Sequence 24, Appl
3	1062.5	23.8	868	4	US-09-398-395A-56	Sequence 56, Appl
4	1062.5	25.8	868	4	US-09-887-586A-56	Sequence 56, Appl
5	1062.5	25.8	868	4	US-09-895-752-56	Sequence 56, Appl
6	1062.5	25.8	868	4	US-09-903-012B-56	Sequence 56, Appl
7	1062.5	25.8	868	4	US-09-900-797-56	Sequence 56, Appl
8	1013.5	24.6	862	3	US-09-315-861-2	Sequence 2, Appli
9	1013.5	24.6	862	4	US-09-398-395A-44	Sequence 44, Appl
10	1013.5	24.6	862	4	US-09-887-586A-44	Sequence 44, Appl
11	1013.5	24.6	862	4	US-09-895-752-44	Sequence 44, Appl
12	1013.5	24.6	862	4	US-09-903-012B-44	Sequence 44, Appl
13	1013.5	24.6	862	4	US-09-593-253-2	Sequence 2, Appli
14	1013.5	24.6	862	4	US-09-900-797-44	Sequence 44, Appl
15	965	23.5	782	4	US-09-360-545-16	Sequence 16, Appl
16	965	23.5	782	4	US-09-398-395A-46	Sequence 46, Appl
17	965	23.5	782	4	US-09-887-586A-46	Sequence 46, Appl
18	965	23.5	782	4	US-09-895-752-46	Sequence 46, Appl
19	965	23.5	782	4	US-09-903-012B-46	Sequence 46, Appl
20	965	23.5	782	4	US-09-900-797-46	Sequence 46, Appl
21	965	23.5	817	3	US-09-234-393-13	Sequence 13, Appl
22	965	23.5	817	3	US-09-234-393-38	Sequence 38, Appl
23	965	23.5	817	3	US-09-234-393-40	Sequence 40, Appl
24	965	23.5	817	4	US-09-865-171-13	Sequence 13, Appl
25	965	23.5	817	4	US-09-865-171-38	Sequence 38, Appl
26	965	23.5	817	4	US-09-865-171-40	Sequence 40, Appl
27	962	23.4	817	3	US-09-234-393-42	Sequence 42, Appl

Db 4 SRPTGVARFAASSSSSSASLPPGVDDVTTTTKTGALHFEETKERIKLFDKVELSVSAYD 63
 Qy 65 TAWAVWSPNSPKSPCPPECLNWLINNOQNDGWSGLVNHHTHHNHPLL-KDSLSSTLAC 123
 Db 64 TAWAVWSPNSLQPLFPFCINWVLDSDHAGSGWLL-----HNDQLMKANLLSTLAC 118
 Qy 124 IVALKRWVGEQINQKGLSFIESNLASATEKSPSPGPDIIPLPGGLLEKAKLDINLLSK 183
 Db 119 VLTLRWNIGHMOKALDFIKSNTASATDQORSVPDPDIIIPFGMIEYAKLDINMLPLA 178
 Qy 184 QTDLSMLHKLREOKRCHSNMDG--YLAYISEGLNLYDMWVKYQKNGSVFNSPS 241
 Db 179 PTVNVALVKKLELRSCHNSGKAYLAYVSEGKIQDMDWVWQYQKNGSLFNSPS 238
 Qy 242 ATAAPFINONPCNLNLSLIDKFGNAVTVTPHDLFIRLSMVDTIERLGHSHHPRVEI 301
 Db 239 TTAAPAFMRNDDGCFYLSLLQKFDGVSPTIPLDIYARLHMVDSLQRFGARHPKEI 298
 Qy 302 KNVLDETYRCWDERDQIFMDVVTTCALAPRLLRINGEVSPDPLAETITNEL-----AL 354
 Db 299 RSVLDETYRCWQGEENIFLDASTCAMAFMRUKRVEGVDVSSDQLTQFSEDIFPNCILGGYL 358
 Qy 355 KDEYAALEYTHASHILYQEDLSGKILK-SADFLKEIISTD---SNRLSKLHKEVENA 410
 Db 359 KDFGASLELYKASQIITHPDESIVENINSWTSRFLKHGLSSDSVWSDRDTSVVKQAVNA 418
 Qy 411 LKPPINTGLERINRNIQIYVNDTRILKTTTHSSNINISWTDYLRILAVEDFYTCQSIYRE 470
 Db 419 LEPPYNATLERLITSKAMESYSGDIIVRISKSPYACILFQGHQDFLELAVEDFNTLRIHLK 478
 Qy 471 ELKGLERWVVENKLDOLKPARQKTAYCYFVSVAATLSPSELSDARISWAKNGIITTVVDDF 530
 Db 479 EUELORWVVENKLDLDELKFFRLHLCYCPAAAATLTDPELHDARIAWAQVLTITVDDF 538
 Qy 531 FDTGGTIDETNLIOQCEKWNVDKDCSEHVRILFLALKAICWIGDFAFKWQARDVT 590
 Db 539 YDGGSEELDNLIELVEKWDPDGEGVYCKDVEIVFLALHSTVCEIGRRALVWQGRSVM 598
 Qy 591 SHVIOQWLEIMNSLREAIWTRDAYVPTLNEYMENAVSPALGPIVKPAIYFVGPKLSEE 650
 Db 599 RNVIDGWLALLKVRKEAEWSTNKNVSPSMGEYMEQARVSPALGPILPLFVFPKPLSEE 658
 Qy 651 IVESSEYHNLFLKMSQGRLLNDIHSFKREFKEKKNALVALHLSNGBSGKVEBEVREMM 710
 Db 659 MIGSEYQKYLKMSLAGRLKNDIRSVDRECKEKNLILSLWIDGGGVTKBEAIBAK 718
 Qy 711 MNIKRRKELMKLIFBENGSIIVPRACKDAFWMCHVLNFFYANDGFTGNTLDTVKDII 770
 Db 719 GDFERAIRELLGLVQEN-TTIPRACKDLFWKLSIVNLFYMEDDGYTSNRLAMTVKAMF 777
 Qy 771 YNPLVL 776
 Db 778 EQPMDL 783

RESULT 2

US-09-614-912-24
 ; Sequence 24, Application US/09614912
 ; Patent No. 6677502
 ; GENERAL INFORMATION:
 ; APPLICANT: Allen, Steve
 ; APPLICANT: Rafaleski, Antoni
 ; APPLICANT: Ozozco, Buddy
 ; APPLICANT: Miao, Gou-Hau
 ; APPLICANT: Famodu, Omolayo O.
 ; APPLICANT: Lee, Jian Ming
 ; APPLICANT: Sakai, Hajime
 ; APPLICANT: Wang, Zude
 ; APPLICANT: Calmi, Perry G
 ; APPLICANT: Anderson, Shawn
 ; TITLE OF INVENTION: Plant Metabolism Genes
 ; FILE REFERENCE: BB1378 US NA
 ; CURRENT APPLICATION NUMBER: US/09/614,912

; CURRENT FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: 60/143,401
 ; PRIOR FILING DATE: 1999-07-12
 ; PRIOR APPLICATION NUMBER: 60/143,412
 ; PRIOR FILING DATE: 1999-07-12
 ; PRIOR APPLICATION NUMBER: 60/146,650
 ; PRIOR FILING DATE: 1999-07-30
 ; PRIOR APPLICATION NUMBER: 60/170,906
 ; PRIOR FILING DATE: 1999-12-15
 ; PRIOR APPLICATION NUMBER: 60/172,959
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/172,946
 ; PRIOR FILING DATE: 1999-12-21
 ; NUMBER OF SEQ ID NOS: 204
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 24
 ; LENGTH: 732
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: (349)
 ; US-09-614-912-24
 Query Match 37.0%; Score 1523; DB 4; Length 732;
 Best Local Similarity 43.7%; Pred. No. 3.1e-126;
 Matches 317; Conservative 133; Mismatches 233; Indels 42; Gaps 8;
 Qy 90 INNOLNDGWSGLVNHHTHHNHPLLKDSLSSTLACIVALKRWVGEQINQKGLSFIESNLA 149
 Db 4 LQNOQDDGSGWV---SQSDSSVSVDVLLSTLACVILKRWVGRNINRWGLHFIENFS 59
 Qy 150 SAPEKSQSPGPIGDIIFPGGLLEYAKNLDINLLSKQTDLSMLHKLREOKRCHSNMDG- 208
 Db 60 VAMDEQFTPIGFENFTFGGLSLGIDMGLEFPVQDVCGLHRRMELKRLAVDSSFG 119
 Qy 209 --YLAYISEGLNLYDMWVKYQKNGSVFNSPSATAAFINQHPGCIYINLSLDKF 266
 Db 120 KATMAFIPGFGNMDWDQVMKFORKNGSLFSTPSTTAVALIHKYNDQALQYLLNLLVNEF 179
 Qy 267 GNAVPTVYPHDLFIRLSMVDTIERLGHSHHFRVEIKNLDETYRCWVERDEQIFMDVVT 326
 Db 180 GSAVPAMPYPSRVHCQSLMVDALKEKGLSQRFVSEISILDMATNCMLONDEELMDIATF 239
 Qy 327 ALAFRLLRINGEVSPDPLAEITNELA-----LKDEYAALEYTHASHI-LYQEDLS 377
 Db 240 AMAFRLLRNGYDVSDDELSHVAGASTFHDLSQGLYNDTKSLLELYKTSKVTLSNDLIL 299
 Qy 378 GKQILKSADFLKEIISTDSNRLSKLHKK-----EVENALKFP 414
 Db 300 DRIGSWSGNLLKDKMCKSVQKTRFLERCKCKQNLNSHFLGSGVYVCVLAQFIEYAVNFP 359
 Qy 415 INTGLERINRNIQIYVNDTRILKTTTHSSNINISWTDYLRILAVEDFYTCQSIYREELKG 474
 Db 360 LYSTLELEKRNIEHFDWAGSLML-TKSSSPRINOEFLLALAVEDFSFORVTRDELRH 418
 Qy 475 LERWVVENKLDOLKPARQKTAYCYFVSVAATLSPSELSDARISWAKNGIITTVDDDFD 534
 Db 419 LDSWVENKLDOLQPARQKLTICYLSAAIVFSSELSDARISWAKNGVLTITVDDDFDVG 478
 Qy 535 GTIDELTNLIQCEKWNVDKDCSEHVRILFLALKAICWIGDFAFKWQARDVTSHVI 594
 Db 479 GSKEELENLIALVEKHGHAHVEFYSEQVIVFSAYITVTVNHLGAMASAAQGRDLTNHLV 538
 Qy 595 QTWLEIMNSLREAIWTRDAYVPTLNEYMENAVSPALGPIVKPAIYFVGPKLSEEVES 654
 Db 539 EIMLDLLRSMMVVEAEWORCOQVPTVEEYIMNAVVSFALGIVLPLALFVGGELLEHA 598
 Qy 655 SEYHNLFKLMSTQGRLLNDIHSFKREFKEKKNALVALHLSNGESGKVEEVEVEMMMK 714
 Db 599 EYDKLFRLYSTCGRLNDYQSLEREGQKGLNSVLSLLVHSGGSMSTRAAKAKMQKSID 658
 Qy 715 NKRKELMKLIFEENGSIIVPRACKDAFWMCHVLNFFYANDGFTG-NTILDTVKDIIYNP 773

```
Db 659 VSRDRLRLVLRKE-SAVERPCKELFKWCKILHLFVSQNDGFSPEKMWVSANVINEP 717
QY 774 LVLVN 778
Db 718 LKVN 722

RESULT 3
US-09-398-395A-56
; Sequence 56, Application US/09398395A
; Patent No. 6468772
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 64687721, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/398,395A
; CURRENT FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/100,993
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 868
; TYPE: PRT
; ORGANISM: Abies grandis
US-09-398-395A-56

Query Match 25.8%; Score 1062.5; DB 4; Length 868;
Best Local Similarity 32.3%; Pred. No. 3.5e-85;
Matches 261; Conservative 148; Mismatches 339; Indels 61; Gaps 17;

QY 7 IASPLLTGKSNRPAAL---SAIHTASTSHGGGTNPINLIIDTTERIKERIQKQKXNV---EISV 60
Db 79 LSSSTLVKREFPPGFWKDDLDLSITSSHKVAASDEKR-IETLISEIKNFRMCGYETNP 137
QY 61 SSYDTAWAMVPSPNSPKSPCFPECLNWLINQLNDGSGVLNHNHHPKLLKDSLSST 120
Db 138 SAYDTAWARIPAVGSDNPHFETVEMILQNLKXGSGW-----EGFYFLAYDRILAT 191
QY 121 LACIVALKRWNVGSDQKNGLSFIESNLASATEKQSP-SPIGFDIIFPGLLEYAKNLDIN 179
Db 192 LACITTLTWRGTQVQKGIETFRTOAGKMEDEADSHRPSGFETVFPFAMLKEAKILGID 251
QY 180 LLSKQTDPSLMLHKELEKQKCHSN---EMDGYLAVISEGLNLYDMNWKYQKNGSV 236
Db 252 LPYDLPFLKQIIEKREAKRIPTDVLVLPALPTLLYSLEGQEIYVDQKIMKLOKSGSF 311
QY 237 FNSPSTATAAFINHQNPCLNLYNSLLDKFGNAVPTVPHDIFIRLSMVDITIERLGISH 296
Db 312 LSSPASTAAVEMRTGKCKCLDFNLVKKFGNHVECHYPDLDFELWAVDIVERLIGDRH 371
QY 297 FRVEIKNVLDITYRCWVERD-----EQIFMDVVTCALAFRLIRINGEYSPDPLAEITNE 351
Db 372 FKEIEKALDYVSHWDERGIGWARENPVPDIDDTAMGLRILRLHGVNVSVDLTKTFRDE 431
QY 352 LALKDEYAA-----LETHASHILYQEDLSGKQILKSA-----DFLKE 390
Db 432 ---NGEFCFCFGLQGTORGVTMDLNVNRCSHVSF-----PGEIMEPAKLCIRYENALEN 483
QY 391 IISTDSNRSLKLIHKEVENALKFPINTGLERINTRRNIOQVNDTRILKTTIYHSSNLSN 450
Db 484 VDAFDKFAFKNIRGEVEYALKYPWPKSMRPLEARSYIENGPDDVWLCKIVYMPYILSN 543
QY 451 TDYLRVAVEDFYTCOSIVREBELKGLERWVVENKLDQLKFAKQTAICYFSVAATLSSPEL 510
```

```
Db 544 EKYLELAKLDENKVKQSIHQTELQDLRRWKKSSGFTDLNTRERVTIYSPSPASFIPEPF 603
QY 511 SDARISWAKNGILTTVVDDFDIGGIDELTNLIQCVERKNVDVDKDCSEHVRILFLAL 570
Db 604 SKREVIYTKTSNFTVILDDLYDAHGSLLDKLFTESVKRWDLSL-VDQMPQOMKICFVG 662
QY 571 KDAICWIGDEAFKQWQARDVTSVHIQTWLELMSMLREALWTRDAYVPTILNEYMENAYVSF 630
Db 663 YNTFDIAKEGERQGRDVLGIQNVKQVQLRAYTKEAEWSEAKYVPSFNEYIENASVSI 722
QY 631 ALGPVIVKPAIYFVGPKLSEIYESSSEYHNLF-KLMSSTOGRLLNDIHSFKREKFKGKLN 689
Db 723 ALGTVVLISALFTGEVLTDEVLSKIDRESRFIQLMGLTGLVNDTXYQAEKQSGV-AS 781
QY 690 ALHLSNGESCKV-EEVEVEEMMMKINKKELMKLIFENGSIIVPRACKDAFWMCMCHVLN 748
Db 782 AIQCYMKDHPKISEEALQHVYSVMENALEELNR---EFVNKIPDIYKRLVFPETARIMQ 838
QY 749 FFYANDDGFT---GNTILDTVKDIIVNPL 774
Db 839 LFYMQGDGLTSLHDMWEIKEHVXNCLFPQV 867

RESULT 4
US-09-887-586A-56
; Sequence 56, Application US/09887586A
; Patent No. 6495354
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 64953541, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/887,586A
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 09/398,395
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 868
; TYPE: PRT
; ORGANISM: Abies grandis
US-09-887-586A-56

Query Match 25.8%; Score 1062.5; DB 4; Length 868;
Best Local Similarity 32.3%; Pred. No. 3.5e-85;
Matches 261; Conservative 148; Mismatches 339; Indels 61; Gaps 17;

QY 7 IASPLLTGKSNRPAAL---SAIHTASTSHGGGTNPINLIIDTTERIKERIQKQKXNV---EISV 60
Db 79 LSSSTLVKREFPPGFWKDDLDLSITSSHKVAASDEKR-IETLISEIKNFRMCGYETNP 137
QY 61 SSYDTAWAMVPSPNSPKSPCFPECLNWLINQLNDGSGVLNHNHHPKLLKDSLSST 120
Db 138 SAYDTAWARIPAVGSDNPHFETVEMILQNLKXGSGW-----EGFYFLAYDRILAT 191
QY 121 LACIVALKRWNVGSDQKNGLSFIESNLASATEKQSP-SPIGFDIIFPGLLEYAKNLDIN 179
Db 192 LACITTLTWRGTQVQKGIETFRTOAGKMEDEADSHRPSGFETVFPFAMLKEAKILGID 251
QY 180 LLSKQTDPSLMLHKELEKQKCHSN---EMDGYLAVISEGLNLYDMNWKYQKNGSV 236
Db 252 LPYDLPFLKQIIEKREAKRIPTDVLVLPALPTLLYSLEGQEIYVDQKIMKLOKSGSF 311
QY 237 FNSPSTATAAFINHQNPCLNLYNSLLDKFGNAVPTVPHDIFIRLSMVDITIERLGISH 296
Db 312 LSSPASTAAVEMRTGKCKCLDFNLVKKFGNHVECHYPDLDFELWAVDIVERLIGDRH 371
```



```
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 868
; TYPE: PRT
; ORGANISM: Abies grandis
US-09-903-012B-56

Query Match      25.8%; Score 1062.5; DB 4; Length 868;
Best Local Similarity 32.3%; Pred. No. 3.5e-85;
Matches 261; Conservative 148; Mismatches 339; Indels 61; Gaps 17;

QY 7 IASPLITKSNRPAAL---SAHTASTSHGGQTNPNTNLIIDTTKERIQOKFKNV---EISV 60
Db LSSSTLVKREPPGFWKDDLDLSLSSHVAASDEKR-IETLISEIKMFCMCGYETNP 137
QY 61 SSYDTAWAVMPSNPKSPCEPCINLWLNQNDGSLVNHNNHHPHLLKDSLSST 120
Db LSSSTLVKREPPGFWKDDLDLSLSSHVAASDEKR-IETLISEIKMFCMCGYETNP 137
QY 138 SAYDTAWAVARIPAVDGSNDPFPETVEMILQNLQKDGSGW-----EGFYFLAYDRILAT 191
QY 121 LACIVALKRWNVGEOINKGLSFIESNLASATEKSQP-SPIGFDIIFPGLLEYAKNLIN 179
Db LACITLTLWTGTQVQKGIEFFRTQAGKMEADSHRPSGFIEVFPAMLEAKILGLD 251
QY 180 LLSKQTDPSLMLHRELEQKCHSN---EMDGYLAYISEGLNLYDMNWKYQMKNGSV 236
Db LPYDLPFLKQIIEKREAKLRIPDVLVYALPTLLYSLGLEQIIVDWQKIMKLOSKDSGF 311
QY 237 FNSPATAAFAINHQNPCLNVLNSLLDKFGNAVPTVPHDLFIRLSMVDTIERLIGISHH 296
Db LSSPASTAAVFMRTGNKKCLDFNLVKKFGNHVPCHYPDLFERLWAVDTVERLIGIDRH 371
QY 297 PRVEIKNVLDETYRCWVERD-----EQIFMDVVTTCALAFRLIRINGEYVSPDLAEITNE 351
Db FKEELKEALDVIYSHWDERGIGWARENVPDIDDTAMGLRIILRHGYNVSSDVLKTFRDE 431
QY 352 LALKDEYAAA-----LETYHASHILYQEDLSGKQILKSA-----DFLKE 390
Db ---NGEPFCFLGQTQGVTDMLNVNRCSHVSP-----PGETIMEEAKLCTERYLENALEN 483
QY 391 IISTDSNRLSLKIHEVENALKFPIGTGLERINTRNLIQNVNDNTRILKTYHSSNISN 450
Db VDAFDKWAFFKNIRGEVEYALKYPWHKSMRLEARSYIENYGPDDVWLKGTVMMPYISN 543
QY 451 TDYLELAVEDFTYCQSIYREELKGLERWVENKLDQKFAKQATACVFSVAATLSSPEL 510
Db EKYLELAKLDFNKVQSIHQTELQDLRRWKKSGGFTDLNFTREVTETIYFSPASFIFEPEP 603
QY 511 SDARISWAKNGILITVVDFFDIGGTIDELTNLIQCVKEKNVVDVKDCCSHVRLIFLAL 570
Db SKREVTYKTSNFTVILDDLYDAHGSLLDLKLFIESVKRWDLSL-VDQMPQOMKICFVGF 662
QY 571 KDAICWIGDEAFKQARQVDTSHVITQWLELNMNMLREAITRDAYVPTLNYMENAYVSF 630
Db YNTFNDIAKEGRERQGRDVLGYIQNVWKVQLEAYTKEAWESEAKYVPSFNEYIENASVSI 722
QY 631 ALGPVIRPAIVFGPKLSEEVESSEVHNLF-KLMSTQGRLLNDIHSFKFEKGLNAV 689
Db ALGTVVLLISALFTGEVLDTVELSKIDRESRLQLMGLTGRLVNDTKTYQAGRGQGEV-AS 781
QY 690 ALHISNGSGKV-EEVVEEEMMMIKKRKEIMKLIIFEENGSIIVPRACKDAFWNMCHVLN 748
Db 782 AIQCVKMDHPKISEEALQHVSVMENALEELNR---EFVANKIPDIYKLVFETARIMQ 838
QY 749 FFYANDDGFT---GNTLDTVKDIYNPL 774
Db 839 LFYMQGDGLTILSHDMEIKERHVKNCILFQPV 867
```

RESULT 7

US-09-900-797-56

; Sequence 56, Application US/09900797

```
; Patent No. 6645762
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 66457621, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/900,797
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/398,395
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 868
; TYPE: PRT
; ORGANISM: Abies grandis
US-09-900-797-56
```

Query Match 25.8%; Score 1062.5; DB 4; Length 868;

Best Local Similarity 32.3%; Pred. No. 3.5e-85;

Matches 261; Conservative 148; Mismatches 339; Indels 61; Gaps 17;

```
QY 7 IASPLITKSNRPAAL---SAHTASTSHGGQTNPNTNLIIDTTKERIQOKFKNV---EISV 60
Db LSSSTLVKREPPGFWKDDLDLSLSSHVAASDEKR-IETLISEIKMFCMCGYETNP 137
QY 61 SSYDTAWAVMPSNPKSPCEPCINLWLNQNDGSLVNHNNHHPHLLKDSLSST 120
Db LSSSTLVKREPPGFWKDDLDLSLSSHVAASDEKR-IETLISEIKMFCMCGYETNP 137
QY 138 SAYDTAWAVARIPAVDGSNDPFPETVEMILQNLQKDGSGW-----EGFYFLAYDRILAT 191
QY 121 LACIVALKRWNVGEOINKGLSFIESNLASATEKSQP-SPIGFDIIFPGLLEYAKNLIN 179
Db LACITLTLWTGTQVQKGIEFFRTQAGKMEADSHRPSGFIEVFPAMLEAKILGLD 251
QY 180 LLSKQTDPSLMLHRELEQKCHSN---EMDGYLAYISEGLNLYDMNWKYQMKNGSV 236
Db LPYDLPFLKQIIEKREAKLRIPDVLVYALPTLLYSLGLEQIIVDWQKIMKLOSKDSGF 311
QY 237 FNSPATAAFAINHQNPCLNVLNSLLDKFGNAVPTVPHDLFIRLSMVDTIERLIGISHH 296
Db LSSPASTAAVFMRTGNKKCLDFNLVKKFGNHVPCHYPDLFERLWAVDTVERLIGIDRH 371
QY 297 PRVEIKNVLDETYRCWVERD-----EQIFMDVVTTCALAFRLIRINGEYVSPDLAEITNE 351
Db FKEELKEALDVIYSHWDERGIGWARENVPDIDDTAMGLRIILRHGYNVSSDVLKTFRDE 431
QY 352 LALKDEYAAA-----LETYHASHILYQEDLSGKQILKSA-----DFLKE 390
Db ---NGEPFCFLGQTQGVTDMLNVNRCSHVSP-----PGETIMEEAKLCTERYLENALEN 483
QY 391 IISTDSNRLSLKIHEVENALKFPIGTGLERINTRNLIQNVNDNTRILKTYHSSNISN 450
Db VDAFDKWAFFKNIRGEVEYALKYPWHKSMRLEARSYIENYGPDDVWLKGTVMMPYISN 543
QY 451 TDYLELAVEDFTYCQSIYREELKGLERWVENKLDQKFAKQATACVFSVAATLSSPEL 510
Db EKYLELAKLDFNKVQSIHQTELQDLRRWKKSGGFTDLNFTREVTETIYFSPASFIFEPEP 603
QY 511 SDARISWAKNGILITVVDFFDIGGTIDELTNLIQCVKEKNVVDVKDCCSHVRLIFLAL 570
Db SKREVTYKTSNFTVILDDLYDAHGSLLDLKLFIESVKRWDLSL-VDQMPQOMKICFVGF 662
QY 571 KDAICWIGDEAFKQARQVDTSHVITQWLELNMNMLREAITRDAYVPTLNYMENAYVSF 630
Db YNTFNDIAKEGRERQGRDVLGYIQNVWKVQLEAYTKEAWESEAKYVPSFNEYIENASVSI 722
QY 631 ALGPVIRPAIVFGPKLSEEVESSEVHNLF-KLMSTQGRLLNDIHSFKFEKGLNAV 689
```

Db 723 ALGTVLISALTEGVLNDEVLSKIDRESRFQMLGTLGRVNDYKYQAERGQGEV-AS 781
 QY 690 ALHLSNGESGKV-EEVVEEMMMIKKRLMKLIFENGSIYVPRACKDAFNMCHVLN 748
 Db 782 AIQYMKDHPKISEEALQHVYSVMENALEELNR---EFVNNKIPDIYKRLVFTARIMQ 838
 QY 749 FRYANDDGT---GNVILDTVKDIYNPL 774
 Db 839 LFYMQDGLTLGSHDMEIKHVRNCLPQV 867

RESULT 8
 US-09-315-861-2
 ; Sequence 2, Application US/09315861
 ; Patent No. 6114160
 ; GENERAL INFORMATION:
 ; APPLICANT: Rodney B. Croteau, Mark R. Wildung
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TAXOL
 ; TITLE OF INVENTION: BIOSYNTHESIS
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Donald L. Stephens Jr.
 ; STREET: One World Trade Center
 ; STREET: 121 S.W. Salmon Street
 ; STREET: Suite 1600
 ; CITY: Portland
 ; STATE: Oregon
 ; COUNTRY: United States of America
 ; ZIP: 97204
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Disk, 3-1/2 inch
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: MS DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/315,861
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/843,363
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Donald L. Stephens Jr.
 ; REGISTRATION NUMBER: 34,022
 ; REFERENCE/DOCKET NUMBER: 4630-46842/DLS
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (503) 226-7391
 ; TELEFAX: (503) 228-9446
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 862 amino acid residues
 ; TYPE: amino acid
 ; STRANDEDNESS: single stranded
 ; TOPOLOGY: linear
 ; US-09-315-861-2

Query Match 24.6%; Score 1013.5; DB 3; Length 862;
 Best Local Similarity 32.2%; Pred. No. 7.9e-81;
 Matches 241; Conservative 139; Mismatches 316; Indels 53; Gaps 16;
 QY 57 EISVSSYDTAWAVP--SPNSPKSPCFECLNWLNNQNDGSGVLNHNHPLLK 114
 Db 134 DISPAVDTAWAVLATISDGEKFPQALNWFNNQDGSWGIESHFS-----LC 187
 QY 115 DLSLSTLACIVALKRWNVGSDQINKLSGTIESNLASATEKSQSPISGDFIIFGILLYAK 174
 Db 188 DRLLNTTNSVIALSWKVTGHSQVQGAETFAENLRLLNEDELSP-DFQIIFPALLOKAK 246
 QY 175 NLDNLILSKQDFSLMLHKB--LEQKRCNSNEMDGLVLAISEGLNLYDNWNVKQYQMK 232
 Db 247 ALGINPLVDLPFIKYLSTTREATLTVSAADNIPANMLNALEGLEEVIDWKNIRFQSK 306

QY 233 NGSVFNPSATAAFINHQNPGLNLYNLKDFGNNAVETVYPHDLFIRLSMVDFTIERIG 292
 Db 307 DGSFLSSPASTACVLMTGDKCFTFLNNLLKDFGCGVPCWYSIDLLERLSLVNDNIHLG 366
 QY 293 ISHFRVEIKNVLDITYRCWVERD-----EQFMDVVTTCALAFRLLRINGEVSDDPLAE 347
 Db 367 IGRHFKQEIKGALDYVYRHWSEGIWGGRDSLVDPDNTTALGLRTHRMHGYNVSSVLNN 426
 QY 348 ITNELALKDE-----YAALETYHSHILYQEDISGKQILKSAADFLKEII 392
 Db 427 -----FKDENGFRFPSSAQGTHVELRSVNLPRASDLAFDPERAMDARKFAEPLYREAL 480
 QY 393 STDSNRLSKLIHKEVENALKFPINTGLERINTRINQIYNVDNTRILKTYTHSSNISWTD 452
 Db 481 ATKISINTKLF-KEIEYVVEYDWMGSIPLREARSYIDSYDDNVWQKTLYRMPSLNSK 539
 QY 453 YLRAVEDFYTCOSIYREBELKGLRWVENKLDQLKFAQKTAICYFVSVAATLSSPELSD 512
 Db 540 CLEAKLDFNIVQSLHQEELKLLTRWKSQSMADINFTRHVAEVYFSSAT--FEFEYSA 597
 QY 513 ARISWAKNGILTTVVDDPDFDIGTIDELTNLIQCVEKNVDV---DKDCCSEHVRILFLA 569
 Db 598 TRIAFTKIGLQVIFDDMADIATLDELKSFTEGVKWDTSLLHEIPECMQTCFKVWEKL 657
 QY 570 LKDAICWIGDEAPKQWQARDVTSVHIVTWELELNMNMLREAIWTRDAVPTILNEYMENAYVS 629
 Db 658 MEB---VNDVVVQGRDMLAHIRKPWELYNFYQVEREWLEAGYIPTFEYLKTYAIS 713
 QY 630 FALGPVTKAIYFVGPKLSEEVESSEY-HNLFKLMSTQGRLLNDIHSKREFKEG-KLN 687
 Db 714 VGLGPTCTLOPILMLGELVKDDVVEKVHPSNMPELVSLSWRLTNDTKTYQAEKARGQAS 773
 QY 688 AVALHLSNGESGKVEEVEEVEEMMMIKKRLMKLIFENGSIYVPRACKDAFNMCHVL 747
 Db 774 GIACYMKD-NPGATEDEDAIKHICRVVDRAKESFEYFKPSNDI-PWGCKSFIFNLRLCV 831
 QY 748 NFFYANDDGF--TGNITLDTVKDIYNPL 774
 Db 832 QIFYKFDGVIAGIANEBIKDYIRKVIDPI 860

RESULT 9
 US-09-398-395A-44
 ; Sequence 44, Application US/09398395A
 ; Patent No. 6468772
 ; GENERAL INFORMATION:
 ; APPLICANT: Chappell, Joseph
 ; APPLICANT: No. 6468772, Joseph P.
 ; APPLICANT: Starks, Courtney M.
 ; APPLICANT: Manna, Kathleen R.
 ; TITLE OF INVENTION: SYNTHASES
 ; FILE REFERENCE: 07678-025001
 ; CURRENT APPLICATION NUMBER: US/09/398,395A
 ; CURRENT FILING DATE: 1999-09-17
 ; PRIOR APPLICATION NUMBER: 60/100,993
 ; PRIOR FILING DATE: 1998-09-18
 ; PRIOR APPLICATION NUMBER: 60/130,628
 ; PRIOR FILING DATE: 1999-04-22
 ; PRIOR APPLICATION NUMBER: 60/150,262
 ; PRIOR FILING DATE: 1999-08-23
 ; NUMBER OF SEQ ID NOS: 58
 ; SOFTWARE: FASTSEQ for Windows Version 3.0
 ; SEQ ID NO 44
 ; LENGTH: 862
 ; TYPE: PRT
 ; ORGANISM: Taxis brevifolia
 ; US-09-398-395A-44

Query Match 24.6%; Score 1013.5; DB 4; Length 862;
 Best Local Similarity 32.2%; Pred. No. 7.9e-81;
 Matches 241; Conservative 139; Mismatches 316; Indels 53; Gaps 16;
 QY 57 EISVSSYDTAWAVP--SPNSPKSPCFECLNWLNNQNDGSGVLNHNHPLLK 114

```
Db 134 DISPAYDTAWVARLATISSDSEKPRFFOALNWVFNQLODQSGWGIESHFS-----LC 187
QY 115 DLSLSTLACIVALKRWNVGEDQINKGLSTIESNLASATEKSQSPGFDIIPGLLEYAK 174
Db 188 DRLLNTNTSVIALSVWKTGHSQVQOGAFIAENLRLLNEEDLSP-DFQIIFPALLQKAK 246
QY 175 NLDINLLSKQTDPSLMLHKRE--LEQKRCHSNEMDGYLAISEGLNLYDNWVWVKYQMK 232
Db 247 ALGINLPYDLPIKYLSTTREARLTDVSAADNIPANMLNALEGLEEVIDWKNIMRFSQK 306
QY 233 NGSVNPSATAAAFINHQNPCLNLYNSLLDKFGNAVTPVPHDLFIRLSMVDTIERLG 292
Db 307 DGSFLSSPASTACVLMNTGDEKCFITLNLKDFGCVPCWYSIDLRLSLVDNIHLG 366
QY 293 ISHHRVETIKNVLDTYRCWVERD-----EQIFMDVVTICALAFRLRLRINGEVSDDPLAE 347
Db 367 IGRHFKQEIKGALDYVYRHWSERGIGWGRDSLVPDLNTTALGLRTLRLWHGVNVSDDLNN 426
QY 348 ITNELALKDE-----YAALETYHSHILYOEGLSSGKQILKSADFLKEII 392
Db 427 -----FKDENGFRFSSAGQTHVELRSVNLFRASDLAPFDDERAMDDARKFAEPYREAL 480
QY 393 STDNRLSKLIHKEVENALKFPINTGLERINTRNIQLYNVNDTRILKTYHSSNISNTD 452
Db 481 ATKISTNTKLF-KEIYVYVYVWMSIPRLEARSYIDSDDNYVWQKTLYRMPISLSNSK 539
QY 453 YRLAVERFYTCOSIYREBELKGLERWVENKLDQKPAROKTAYCFVVAATLSPPELSD 512
Db 540 CLEAKLDFNIVQSLHQBELKLLTRWKEGSMADINFTRHRAEVYFSSAT--PEPEYSA 597
QY 513 ARISWAKNGILTTVVDDDFDIDGTTIDELTNLIQCEKKNVDV--DKDCCSEHVRILFLA 569
Db 598 TRIATFKIGCQLVFDMDADIFATIDELKSPTEGVKRWDTSLHLHPECMQTCFKVWFKL 657
QY 570 LKDAICWTIGDEAFKQWQARDVTSVHIQTLWELMNSMLREAIWTRDAYVPTLNEYMENAYVS 629
Db 658 MEE-----VNNDVVKVQGRDMLAHIRKPWELYNFCVQEREWLEAGYIPTFBEYLKTYAIS 713
QY 630 FALGPVIVKPAIVFGPKLSEBEEVSESEY-HNLFKLMSTQGRLLNDIHSFKBEFKEG-KLN 687
Db 714 VGLGCTIQLPILLMGEVLKDDVVEKHYPSPNNFELVSLWSRLTNDTKTYQAEKARGOQAS 773
QY 688 AVALHLSNGESGKVEEVEEVEEMMMKKNRKLMLKLIPEENGSIIVPRACKDAFNMCHVL 747
Db 774 GIACVMKD-NFGATEDEAKHICRVVDRALKEASFEYFKPSNDI-PMGCKSFIFNRLCV 831
QY 748 NFFYANDDGF--TGNITLDTVKDIIYNPL 774
Db 832 QIFVKFIDGYGIANBEIKDYIRKVYIDPI 860
```

RESULT 10

```
US-09-887-586A-44
; Sequence 44, Application US/09887586A
; Patent No. 6495354
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 64953541, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/887,586A
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 09/398,395
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 44
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Taxis brevifolia
US-09-887-586A-44
```

Query Match 24.6%; Score 1013.5; DB 4; Length 862;
Best Local Similarity 32.2%; Pred. No. 7.9e-81;

Matches 241; Conservative 139; Mismatches 316; Indels 53; Gaps 16;

```
QY 57 EISVSYSYTAWAMVP--SPNSPKPCPEPCINLNNQNDGSLGVLNHNTHNHPLIK 114
Db 134 DISPAYDTAWVARLATISSDSEKPRFFOALNWVFNQLODQSGWGIESHFS-----LC 187
QY 115 DLSLSTLACIVALKRWNVGEDQINKGLSTIESNLASATEKSQSPGFDIIPGLLEYAK 174
Db 188 DRLLNTNTSVIALSVWKTGHSQVQOGAFIAENLRLLNEEDLSP-DFQIIFPALLQKAK 246
QY 175 NLDINLLSKQTDPSLMLHKRE--LEQKRCHSNEMDGYLAISEGLNLYDNWVWVKYQMK 232
Db 247 ALGINLPYDLPIKYLSTTREARLTDVSAADNIPANMLNALEGLEEVIDWKNIMRFSQK 306
QY 233 NGSVNPSATAAAFINHQNPCLNLYNSLLDKFGNAVTPVPHDLFIRLSMVDTIERLG 292
Db 307 DGSFLSSPASTACVLMNTGDEKCFITLNLKDFGCVPCWYSIDLRLSLVDNIHLG 366
QY 293 ISHHRVETIKNVLDTYRCWVERD-----EQIFMDVVTICALAFRLRLRINGEVSDDPLAE 347
Db 367 IGRHFKQEIKGALDYVYRHWSERGIGWGRDSLVPDLNTTALGLRTLRLWHGVNVSDDLNN 426
QY 348 ITNELALKDE-----YAALETYHSHILYOEGLSSGKQILKSADFLKEII 392
Db 427 -----FKDENGFRFSSAGQTHVELRSVNLFRASDLAPFDDERAMDDARKFAEPYREAL 480
QY 393 STDNRLSKLIHKEVENALKFPINTGLERINTRNIQLYNVNDTRILKTYHSSNISNTD 452
Db 481 ATKISTNTKLF-KEIYVYVYVWMSIPRLEARSYIDSDDNYVWQKTLYRMPISLSNSK 539
QY 453 YRLAVERFYTCOSIYREBELKGLERWVENKLDQKPAROKTAYCFVVAATLSPPELSD 512
Db 540 CLEAKLDFNIVQSLHQBELKLLTRWKEGSMADINFTRHRAEVYFSSAT--PEPEYSA 597
QY 513 ARISWAKNGILTTVVDDDFDIDGTTIDELTNLIQCEKKNVDV--DKDCCSEHVRILFLA 569
Db 598 TRIATFKIGCQLVFDMDADIFATIDELKSPTEGVKRWDTSLHLHPECMQTCFKVWFKL 657
QY 570 LKDAICWTIGDEAFKQWQARDVTSVHIQTLWELMNSMLREAIWTRDAYVPTLNEYMENAYVS 629
Db 658 MEE-----VNNDVVKVQGRDMLAHIRKPWELYNFCVQEREWLEAGYIPTFBEYLKTYAIS 713
QY 630 FALGPVIVKPAIVFGPKLSEBEEVSESEY-HNLFKLMSTQGRLLNDIHSFKBEFKEG-KLN 687
Db 714 VGLGCTIQLPILLMGEVLKDDVVEKHYPSPNNFELVSLWSRLTNDTKTYQAEKARGOQAS 773
QY 688 AVALHLSNGESGKVEEVEEVEEMMMKKNRKLMLKLIPEENGSIIVPRACKDAFNMCHVL 747
Db 774 GIACVMKD-NFGATEDEAKHICRVVDRALKEASFEYFKPSNDI-PMGCKSFIFNRLCV 831
QY 748 NFFYANDDGF--TGNITLDTVKDIIYNPL 774
Db 832 QIFVKFIDGYGIANBEIKDYIRKVYIDPI 860
```

RESULT 11

```
US-09-895-752-44
; Sequence 44, Application US/09895752
; Patent No. 6559297
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 65592971, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
```

```
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/895,752
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/398,395
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/100,993
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Taxus brevifolia
US-09-895-752-44

Query Match      24.6%; Score 1013.5; DB 4; Length 862;
Best Local Similarity 32.2%; Pred. No. 7.9e-81;
Matches 241; Conservative 139; Mismatches 316; Indels 53; Gaps 16;

QY 57 EISVSSYDTAWAMVP--SPNSPKSPCFPECLNWLNNQDGLVNHNNHPLK 114
Db 134 DISPSAYDTAWARLATISSDGSEKPRFPQALNWFNNQDQSGWGIESHFS-----LC 187
QY 115 DLSSTLACIVALKRWNVGDDQINKGLSFIESNLASATEKSQSPGFIIFPGILEYAK 174
Db 188 DRLNTNSVIALSVKMGTHSQVQQAETAEENLRLLNEDELSF-DFOIIFPALLQKAK 246
QY 175 NLDINLLSKQTDPSMLHKE--LEQKRCHSNEMDGYLAIYISGLGNLYDNNMVKYQMK 232
Db 247 ALGINLPDLPFTKYLSTTREARLTDVSAADNIPANMLNALLGLEBEVDNWKIMRFSQK 306
QY 233 NGSVFNSPATAAFAFINHQNPGLNVLNLLDKFGNAVPTVYPHDLFIRLSMVDTTIERLG 292
Db 307 DGSFLLSPASTACVLMNTGDEKFTFLNLLDKFGGVCPCWYSDILRLSLVDNIEHLG 366
QY 293 ISHFRVEIKNVLDETVRCWVERD-----EQIFMDVVTCALAFRLARINGEYSPDPLAE 347
Db 367 IGRHFQKEIKGALDYVYRHSERIGWGRDLSVPDLNTTALGLRTLRLMHGYNVSSDVLNN 426
QY 348 ITNELAKDB-----YAALETYHSHLYQEDLSSGKQILKSADFLKEII 392
Db 427 -----FKDENGFRFFSAGQTHVELRSVNLFRASDLAFDPERAMDDARKFAEPYLRAL 480
QY 393 STDSNRLSLKHKEVENALKFPINTGLERINTRNQOLYNVDNTRILKTYTHSSNINTD 452
Db 481 ATKISTNTKLF-KEIEYVVEYPWHSIPRLARSYIDSYDNNYVWQKTYLRMPSLNSK 539
QY 453 YRLAIVEDFTYCOSIYREELKGLERWVVENKLDQKFAKQKATAYCFPSVAATLSSPELSD 512
Db 540 CLELAKLDFNIVQSLHQEELKLLTRWKESGMADINFTHRVAEVYFSSAT--FEPEYSA 597
QY 513 ARISWAKNGILTTVDDDFDGGTIDELTNLOCVKEKNVDV---DKDCCESHVRILFLA 569
Db 598 TRAFYKIGLQVLFDMDAIFATLDKLSFTGKVDTSLLHEIPECMTQCFKWWFKL 657
QY 570 LKDAICWIGDEAFKQKQADVSHVIOTWLELMNSMLREAIWTRDAYVPTTLNENYENAYS 629
Db 658 MEE-----VNNDVVKVQGRDMLAHIRKPWELYNFCYQEREWLEAGYIPTPEEYLKTYAIS 713
QY 630 FALGPVVKPAIYFGPKLSIEVESSEY-HNLFKLMSTQGRLLNDIHSFKPEFKG-KLN 687
Db 714 VGLGPCVTLQPIILMGLVKKDDVKEKHYPSNMFELSVLSRLNTDITKTYQAEKARGOQAS 773
QY 688 AVALHLSNGSGKVEEVEEVEEMMMIKKKEKELMKLIFEEBNGSIVPRACKDAFNMCHVL 747
Db 774 GIACYMKD-NPGATEEDAIRKICRVVDRLKESAFEFYKPSNDI-PWGCKSFIFNLCLV 831
QY 748 NFFVANDGPR--TGNTILTDTVKLIYNPL 774
```

```
Db 832 QIFYKFDGCIANBEIKDYIRKVIDPI 860

RESULT 12
US-09-903-012B-44
; Sequence 44, Application US/09903012B
; Patent No. 6569656
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 65696561, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/903,012B
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 60/100,993
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Taxus brevifolia
US-09-903-012B-44

Query Match      24.6%; Score 1013.5; DB 4; Length 862;
Best Local Similarity 32.2%; Pred. No. 7.9e-81;
Matches 241; Conservative 139; Mismatches 316; Indels 53; Gaps 16;

QY 57 EISVSSYDTAWAMVP--SPNSPKSPCFPECLNWLNNQDGLVNHNNHPLK 114
Db 134 DISPSAYDTAWARLATISSDGSEKPRFPQALNWFNNQDQSGWGIESHFS-----LC 187
QY 115 DLSSTLACIVALKRWNVGDDQINKGLSFIESNLASATEKSQSPGFIIFPGILEYAK 174
Db 188 DRLNTNSVIALSVKMGTHSQVQQAETAEENLRLLNEDELSF-DFOIIFPALLQKAK 246
QY 175 NLDINLLSKQTDPSMLHKE--LEQKRCHSNEMDGYLAIYISGLGNLYDNNMVKYQMK 232
Db 247 ALGINLPDLPFTKYLSTTREARLTDVSAADNIPANMLNALLGLEBEVDNWKIMRFSQK 306
QY 233 NGSVFNSPATAAFAFINHQNPGLNVLNLLDKFGNAVPTVYPHDLFIRLSMVDTTIERLG 292
Db 307 DGSFLLSPASTACVLMNTGDEKFTFLNLLDKFGGVCPCWYSDILRLSLVDNIEHLG 366
QY 293 ISHFRVEIKNVLDETVRCWVERD-----EQIFMDVVTCALAFRLARINGEYSPDPLAE 347
Db 367 IGRHFQKEIKGALDYVYRHSERIGWGRDLSVPDLNTTALGLRTLRLMHGYNVSSDVLNN 426
QY 348 ITNELAKDB-----YAALETYHSHLYQEDLSSGKQILKSADFLKEII 392
Db 427 -----FKDENGFRFFSAGQTHVELRSVNLFRASDLAFDPERAMDDARKFAEPYLRAL 480
QY 393 STDSNRLSLKHKEVENALKFPINTGLERINTRNQOLYNVDNTRILKTYTHSSNINTD 452
Db 481 ATKISTNTKLF-KEIEYVVEYPWHSIPRLARSYIDSYDNNYVWQKTYLRMPSLNSK 539
QY 453 YRLAIVEDFTYCOSIYREELKGLERWVVENKLDQKFAKQKATAYCFPSVAATLSSPELSD 512
Db 540 CLELAKLDFNIVQSLHQEELKLLTRWKESGMADINFTHRVAEVYFSSAT--FEPEYSA 597
QY 513 ARISWAKNGILTTVDDDFDGGTIDELTNLOCVKEKNVDV---DKDCCESHVRILFLA 569
Db 598 TRAFYKIGLQVLFDMDAIFATLDKLSFTGKVDTSLLHEIPECMTQCFKWWFKL 657
QY 570 LKDAICWIGDEAFKQKQADVSHVIOTWLELMNSMLREAIWTRDAYVPTTLNENYENAYS 629
Db 658 MEE-----VNNDVVKVQGRDMLAHIRKPWELYNFCYQEREWLEAGYIPTPEEYLKTYAIS 713
```

QY 630 FALGPVPAIYFVGPKLSEIEVESSEY-HNLFKLMSTQRLNDIHSFKRFEKG-KLN 687
D6 714 VGLGCTLOPILLMGLVKDDVEKVVHPSNMFVLSLWRLTNDTKITFOAKARQQAS 773
QY 688 AVALHLSNGESKVEEVVEENMMKKNRKBELMKLIFEENGSIYPRACKDAFNMCHYL 747
D6 774 GIACYMKD-NPGATBEDAIKHICRVVDRAKLEASFEYFKPSNDI-PMGCKSFIFNLRLCV 831
QY 748 NFFYANDDGF--TGNLTLDTVKDIYNPL 774
D6 832 QIFYKFDGYGIANEIKDYIRKVVIDPI 860

RESULT 13
US-09-593-253-2
; Sequence 2, Application US/09593253
; Patent No. 6610527
; GENERAL INFORMATION:
; APPLICANT: Rodney B. Croteau, Mark R. Wildung
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TAXOL
; BIOSYNTHESIS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Donald L. Stephens Jr.
; STREET: One World Trade Center
; 121 S.W. Salmon Street
; Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/593,253
; FILING DATE: 13-Jun-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/843,363
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Donald L. Stephens Jr.
; REGISTRATION NUMBER: 34,022
; REFERENCE/DOCKET NUMBER: 4630-46842/DLS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (503) 226-7391
; TELEFAX: (503) 228-9446
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 862 amino acid residues
; TYPE: amino acid
; STRANDEDNESS: single stranded
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-593-253-2

Query Match 24.6%; Score 1013.5; DB 4; Length 862;
Best Local Similarity 32.2%; Pred. No. 7.9e-81;
Matches 241; Conservative 139; Mismatches 316; Indels 53; Gaps 16;

QY 57 EISVSSYDTAWAMP--SPNSPKSPCEPCINLNNQNDGSLVNNHNNHPLK 114
D6 134 DISPAYTAWARLATISSDGEKPRFPQALNMFNNQLODQSGWGESHS-----LC 187
QY 115 DLSLSLACIVALKRWNVGEQINKGLFIBSNLASATEKSQSPSIGFDIIFPGILEYAK 174
D6 188 DRLNTNNSIALSVSWKTHGSHVQVQGAFFIENURLNNEDELSF-DQIIFPALLQAK 246
QY 175 NLDINLASKOTDFSIMLHKRE--LEQKRCNSNEMDGYLAIYISGLGNLYDNMVKYQMK 232

D6 247 ALGINLPYDLPIKXILSTREARLTDVSAADNI-PANMLNALEGLEEVLDNWKIVRFSQK 306
QY 233 NGSVNSPSATAAFINQNPCLNYSLLDKFGNAVPTVYPHDLFIRLSMVDTIERLG 292
D6 307 DGSFLSSPASTACVLMNTGDEKCFITLNNLLDKFGCVPCMYSIDLLERLSLDVNIHLG 366
QY 293 TSHFRVRIKXVLDIYRCWVERD-----EQIFMDVVTICALAFRLRLRINGYVSPDPLAE 347
D6 367 IGRHFQEIKGALDYVYRHSERGIWGGRDLSVLPDLNTALGLRLTRMHHYVNSDVLNN 426
QY 348 ITNELALKDE-----YAALETYHSHILYQEDLSSGKQLKLSADFLKEII 392
D6 427 -----FKDENGREFSSAGQTHVELRSVNLPRASDLAPPDERAMDARKPAEPIREAL 480
QY 393 STDSNRLSKLIHKEVENALKPPINTGLERINTRNIQLYNVNDNTRILKTYHSSNISNTD 452
D6 481 ATKISTNTKLF-KELEYVVEYFPHMSIPRLEARSYIDSYDDNYVQWQRTLYRMPSLNSK 539
QY 453 YLRLAVEDFYTCOSIYREELKGLERWVVENKLDQLKEAROKTAYCYVPSVAATLSSPELSD 512
D6 540 CLELAKLDFNIVQSLHQEELKLLTRWKESGADINFTRHVAEYVFSAT--FEPEYSA 597
QY 513 ARISWAKNGILTTVVDDFFDIGTIDELTNLIQCVEKNVDY--DKDCCSEHVRILFLA 569
D6 598 TRIAFTKIGLQVLFPDDMADIFATLDELKSTFEGVKRWDTSLHHEIPECWQCFKVFKL 657
QY 570 LKDALCWTGDEAFKQWQARDVTSHTVQWLELMNSMLRAIWTDRDAYVPTLNEYMNAYVS 629
D6 658 MEE-----VNDVVKVQGRDMLAHIRKPMELVFCYVQREMLEAGYIPTFEYLKTYAIS 713
QY 630 FALGPVPAIYFVGPKLSEIEVESSEY-HNLFKLMSTQRLNDIHSFKRFEKG-KLN 687
D6 714 VGLGCTLOPILLMGLVKDDVEKVVHPSNMFVLSLWRLTNDTKITFOAKARQQAS 773
QY 688 AVALHLSNGESKVEEVVEENMMKKNRKBELMKLIFEENGSIYPRACKDAFNMCHYL 747
D6 774 GIACYMKD-NPGATEEDAIIKHICRVVDRAKLEASFEYFKPSNDI-PMGCKSFIFNLRLCV 831
QY 748 NFFYANDDGF--TGNLTLDTVKDIYNPL 774
D6 832 QIFYKFDGYGIANEIKDYIRKVVIDPI 860

RESULT 14
US-09-900-797-44
; Sequence 44, Application US/09900797
; Patent No. 6645762
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 66457621, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/900,797
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/398,395
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Taxus brevifolia
US-09-900-797-44

Query Match 24.6%; Score 1013.5; DB 4; Length 862;
Best Local Similarity 32.2%; Pred. No. 7.9e-81;

Query Match	23.58;	Score 965;	DB 4;	Length 782;
Best local Similarity	31.86;	Pred. No. 1.4e-76;		
Matches 240;	Conservative 145;	Mismatches 318;	Indels 52;	Gaps 15;
Qy	58	ISVSSYDTAWAMVSPNSPKSPCFPCLNWLINQLNDGSGWGLVHNTHHNHPLLKDSL	117	
Db	44	ITPSAYDTAWARVPAIDGSGARQFPQTVDMILKNQLKDGSGWI-----QSHLLSDRL	97	
Qy	118	SSTLACTIVALKRMNVGBDQKNGKLSFTIESNLASATEKSQSPPI--GFDIIPGLLEVAKN	175	
Db	98	LATLSCLVLLKNWVGLQVEQIEFTKSNLELVKDETDQDSLVDTFEIIIPPSLLREAQS	157	
Qy	176	LDINLLSKQTDFSIMLHKRELEQKRCHSNE---MDGYLAVISEGLNLYDWNVMVKYQMK	232	
Db	158	LRGLPVDLPIYHLQTKQERLAKLSREHLYAVSPLLSLKLGQIDIVENERIMEVQSQ	211	
Qy	233	NGSVNFSPSATAAFINHQHPGCLNYLNSLLDKFGNAVPTVYPHDLFIRLSMVDTIERLG	292	
Db	218	DGSLSPSPASTACVFMETGDAKCLEFLNSVMKFGNFVPCLYPVDLLERLLIVDNVRLG	277	
Qy	293	ISHFRVEIKNVLDTEYRCWVERD-----EQIFMDVVTCAAFPLLRINGVEVSP--DP	344	
Db	278	IYRHFEXEIKALDYVYRHHNERIGWGRNLPIADLETTALGFLRLHRYNVSPFAIPDN	337	
Qy	345	LAETITNELAL-----KDEVAALETHASHILYQEDLSSGKOILLKSA-----DPLKEII-	392	
Db	338	FKDANGKPISTGCFNKDKVASMLNLYRASQLAF-----PGENILDEAKSPATKYLREALE	392	
Qy	393	-----STDNRLSKLIIHKEVENALKFPINTGLEERINTERNIQLYNDVNRILKTYTHSSN	447	
Db	393	KSETSSAWNKNQ--LQSEIKYALKTSWHA5VPRVEAKRYCQVYRPDVARIKACVYKLPY	450	
Qy	448	ISNTDYLRLAVEDPYTCOSIYREBLKGLRWVVENKLDOLKPAQKATACVFSVAATLSS	507	
Db	451	VNNEKPLELGLKDFENIIQSHIQEEMKNVTSWFRDSGLPLFTTFARERPLEFYFLVAAGPYE	510	
Qy	508	PELSDAISNAKNGILATVVDPDFDICTIDELNTLTCQVEKNWVDKDCSCSEHRLIF	567	
Db	511	POYAKCRLFTKVAQLQVLDDMDYDTGTLDLKLFTAEVRRWDLSTFN-LPDYMKLCY	569	
Qy	568	LALKDAICWIGDEAFKQARDVTSHVITQWLLELMNSMLREAINTRDAYVPTPLNMEYNAY	627	
Db	570	QIYYDIIVHEVAWEAKEQGRELVFFRKGMEDYLLGYEAEWALAAEYVPTLDEYIKNGI	629	
Qy	628	VSPALGPVVKPAIYFV-GPKLSSEIVSESVYHN--LFLKLMSTOGRLLNDIHSKPEFKE	683	
Db	630	TSIGQRILLSGVLIMDQGLLSQALEKVDYFGRRVITELNSLISRLADDTKYTKAEKAR	689	
Qy	684	GKLNVALHLSNGBSGKVEEVBEMMMIKKKEKELMKLIFEENGSIYVPRACKDAFNMN	743	
Db	690	GELASSIECYMKDPECTEEBALDHIYSLBPVAKELTREFLKPPD--VPPACKKWLFE	747	
Qy	744	CHVLNFFYANDDGF--TGNITLDTVVKOIIYNPIVL	776	
Db	748	TRVTWVIFPKDGDGFGVSKLVEYKDHICEKLIBFLPL	782	

RESULT 15
US-09-360-545-16
; Sequence 16, Application US/09360545
; Patent No. 6429014
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Bohlmann, Jorg
; APPLICANT: Steele, Christopher L
; APPLICANT: Phillips, Michael A
; TITLE OF INVENTION: MONOTERPENE SYNTHASES FROM GRAND FIR (ABIES GRANDIS)
; FILE REFERENCE: wsl13885
; CURRENT APPLICATION NUMBER: US/09/360,545
; CURRENT FILING DATE: 1999-07-26
; EARLIER APPLICATION NUMBER: 60/052,249
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: PCT/US98/14528
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 107

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 08:53:05 ; Search time 73.5657 Seconds
(without alignments)
3445.655 Million cell updates/sec

Title: US-10-041-018-383
Perfect score: 4113
Sequence: 1 MNLSLCIASPLTKSRPAALSAIHTASTSGGQNPNTNLIIDTTKRIQKQKNEIVS 60

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1360919 seqs, 32331874 residues

Total number of hits satisfying chosen parameters: 1360919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PTC_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PTCUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4113	100.0	784	15	US-10-041-018-383
2	4026	97.9	784	15	US-10-041-018-384
3	2855.5	69.4	788	15	US-10-041-018-390
4	2101.5	51.1	785	15	US-10-041-018-389
5	2023.5	49.2	789	15	US-10-041-018-395
6	1675	40.7	780	15	US-10-425-114-65662
7	1537.5	37.4	740	15	US-10-259-194A-108
8	1458	35.4	689	16	US-10-437-963-121564
9	1319	32.1	770	16	US-10-437-963-138526
10	1273	31.0	730	16	US-10-437-963-174639
11	1254	30.5	628	15	US-10-425-114-49083
12	1209	29.4	1157	16	US-10-437-963-174635
13	1075	26.1	795	13	US-10-041-007-37
14	1075	26.1	814	13	US-10-041-007-35

15	1075	26.1	873	13	US-10-041-007-2	Sequence 2, Appli
16	1075	26.1	873	13	US-10-041-007-33	Sequence 33, Appl
17	1075	26.1	873	15	US-10-041-018-398	Sequence 398, App
18	1062.5	25.8	868	9	US-09-895-752-56	Sequence 56, Appl
19	1062.5	25.8	868	9	US-09-887-586A-56	Sequence 56, Appl
20	1062.5	25.8	868	9	US-09-903-012-56	Sequence 56, Appl
21	1062.5	25.8	868	10	US-09-900-797-56	Sequence 56, Appl
22	1062.5	25.8	868	11	US-09-893-820-56	Sequence 56, Appl
23	1062.5	25.8	868	13	US-10-041-007-4	Sequence 4, Appli
24	1062.5	25.8	868	15	US-10-041-018-388	Sequence 388, App
25	1054	25.6	746	13	US-10-041-007-39	Sequence 39, Appl
26	1013.5	24.6	862	9	US-09-895-752-44	Sequence 44, Appl
27	1013.5	24.6	862	9	US-09-887-586A-44	Sequence 44, Appl
28	1013.5	24.6	862	9	US-09-903-012-44	Sequence 44, Appl
29	1013.5	24.6	862	10	US-09-900-797-44	Sequence 44, Appl
30	1013.5	24.6	862	11	US-09-893-820-44	Sequence 44, Appl
31	1013.5	24.6	862	13	US-10-041-007-41	Sequence 41, Appl
32	1013.5	24.6	862	15	US-10-041-018-386	Sequence 386, App
33	965	23.5	782	9	US-09-895-752-46	Sequence 46, Appl
34	965	23.5	782	9	US-09-887-586A-46	Sequence 46, Appl
35	965	23.5	782	9	US-09-903-012-46	Sequence 46, Appl
36	965	23.5	782	10	US-09-900-797-46	Sequence 46, Appl
37	965	23.5	782	11	US-09-893-820-46	Sequence 46, Appl
38	965	23.5	782	13	US-10-041-007-16	Sequence 16, Appl
39	965	23.5	782	14	US-10-025-145A-16	Sequence 16, Appl
40	965	23.5	817	9	US-09-865-171-13	Sequence 13, Appl
41	965	23.5	817	9	US-09-865-171-38	Sequence 38, Appl
42	965	23.5	817	9	US-09-865-171-40	Sequence 40, Appl
43	965	23.5	817	13	US-10-041-007-15	Sequence 15, Appl
44	962	23.4	817	9	US-09-865-171-42	Sequence 42, Appl
45	930	22.6	470	16	US-10-437-963-196878	Sequence 196878,

ALIGNMENTS

RESULT 1

US-10-041-018-383
; Sequence 383, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; PRIOR FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; CURRENT FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 383
; LENGTH: 784
; TYPE: PRT
; ORGANISM: Stevia rebaudiana
US-10-041-018-383

Query Match 100.0%; Score 4113; DB 15; Length 784;
Best Local Similarity 100.0%; Pred. No. 8.9e-298;
Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNLSLCIASPLTKSRPAALSAIHTASTSGGQNPNTNLIIDTTKRIQKQKNEIVS 60

Db 1 MNLSLCIASPLTKSRPAALSAIHTASTSGGQNPNTNLIIDTTKRIQKQKNEIVS 60

QY 61 SSYDTAWAMVPSNPKSPCFPECLNWLNNQNDGSLVNHNNHNNHPLKDSLSST 120

Db 61 SSYDTAWAMVPSNPKSPCFPECLNWLNNQNDGSLVNHNNHNNHPLKDSLSST 120

QY 121 LACIVALKRWNVGDIQNKGLSFIESNLSATEKSPSPIDFIIFPGLLYAKNLDINL 180

Db 121 LACIVALKRWNVGDIQNKGLSFIESNLSATEKSPSPIDFIIFPGLLYAKNLDINL 180

QY 181 LSKQTFSLMLHKELEOKRCHSNEMDGYLAYSGLGNLYDMNMVKYOMKNGSVFNSP 240
DB 181 LSKQTFSLMLHKELEOKRCHSNEMDGYLAYSGLGNLYDMNMVKYOMKNGSVFNSP 240
QY 241 SATAAAFINQHPGCLNLSLDKFGNAVPTVYPHDLFIRLSMVDTIERLGI SHHFRVE 300
DB 241 SATAAAFINQHPGCLNLSLDKFGNAVPTVYPHDLFIRLSMVDTIERLGI SHHFRVE 300
QY 301 IKNVLDTRYCWRVERDEQIFMDVVTTCALAPRLRIRINGEYVSPDPLAETINELAKDEYAA 360
DB 301 IKNVLDTRYCWRVERDEQIFMDVVTTCALAPRLRIRINGEYVSPDPLAETINELAKDEYAA 360
QY 361 LETYHASHILYQEDLSGKQILKASADFLKEIIISTDSNRSLKLIHKEVENALKPPI NTGLE 420
DB 361 LETYHASHILYQEDLSGKQILKASADFLKEIIISTDSNRSLKLIHKEVENALKPPI NTGLE 420
QY 421 RINTRENIQLYNDNTRILKTTYHSSNISNTDYLRILAVEDFYTCQSIYREBELKGLERWV 480
DB 421 RINTRENIQLYNDNTRILKTTYHSSNISNTDYLRILAVEDFYTCQSIYREBELKGLERWV 480
QY 481 ENKLDQKFAKQKTAICYFSVAATLSPELSDARISWAKNGIILTUVDDDFDGGTIDEL 540
DB 481 ENKLDQKFAKQKTAICYFSVAATLSPELSDARISWAKNGIILTUVDDDFDGGTIDEL 540
QY 541 TNLIOQVEKWNVDVDDKCCSEHVRILFLALKDAICWIGDEAFKQWQARDVTSHVITQWLEL 600
DB 541 TNLIOQVEKWNVDVDDKCCSEHVRILFLALKDAICWIGDEAFKQWQARDVTSHVITQWLEL 600
QY 601 MNSMLREAIWTRDAYVPTLNEYMENAYVSPALGPVVKPAIFYVGPGLSEBIVSESYHNL 660
DB 601 MNSMLREAIWTRDAYVPTLNEYMENAYVSPALGPVVKPAIFYVGPGLSEBIVSESYHNL 660
QY 661 FKLMSQGRLLNDIHSFKREFKGLNAVALHLSNGESGKVEBEVVEEMMMIKNKKREL 720
DB 661 FKLMSQGRLLNDIHSFKREFKGLNAVALHLSNGESGKVEBEVVEEMMMIKNKKREL 720
QY 721 MKLIFEENGSI VPRACKDAFWNMCHVNLFFYANDDGTGNTILDTVKDIIYNPLVLNEN 780
DB 721 MKLIFEENGSI VPRACKDAFWNMCHVNLFFYANDDGTGNTILDTVKDIIYNPLVLNEN 780
QY 781 EEQR 784
DB 781 EEQR 784
RESULT 2
US-10-041-018-384
; Sequence 384, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080USI/10025547
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US/10/041.018
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 384
; LENGTH: 784
; TYPE: PRT
; ORGANISM: Stevia rebaudiana
US-10-041-018-384
Query Match 97.9%; Score 4026; DB 15; Length 784;
Best Local Similarity 98.0%; Pred. No. 2.7e-231;
Matches 768; Conservative 8; Mismatches 8; Indels 0; Gaps 0;
QY 1 MNLISCTASPLLTSSNPAISATHTASTSHGGQTNPTNLIDTTKRIOKFOKNVSI 60
DB 1 MNLISCTASPLLTSSNPAISATHTASTSHGGQTNPTNLIDTTKRIOKFOKNVSI 60

QY 61 SSYDTAWVAMPSPNSPKSPCFPECLNWLINQNDGSLVNHHTNHNHPLKDSLSST 120
DB 61 SSYDTAWVAMPSPNSPKSPCFPECLNWLINQNDGSLVNHHTNHNHPLKDSLSST 120
QY 121 LACTIVALKRWNVGEDQINKGLSFIESNLASATEKSQSPIGFDIIPFGLLEYAKNLDINL 180
DB 121 LACTIVALKRWNVGEDQINKGLSFIESNLASATEKSQSPIGFDIIPFGLLEYAKNLDINL 180
QY 181 LSKQTFSLMLHKELEOKRCHSNEMDGYLAYSGLGNLYDMNMVKYOMKNGSVFNSP 240
DB 181 LSKQTFSLMLHKELEOKRCHSNEMDGYLAYSGLGNLYDMNMVKYOMKNGSVFNSP 240
QY 241 SATAAAFINQHPGCLNLSLDKFGNAVPTVYPHDLFIRLSMVDTIERLGI SHHFRVE 300
DB 241 SATAAAFINQHPGCLNLSLDKFGNAVPTVYPHDLFIRLSMVDTIERLGI SHHFRVE 300
QY 301 IKNVLDTRYCWRVERDEQIFMDVVTTCALAPRLRIRINGEYVSPDPLAETINELAKDEYAA 360
DB 301 IKNVLDTRYCWRVERDEQIFMDVVTTCALAPRLRIRINGEYVSPDPLAETINELAKDEYAA 360
QY 361 LETYHASHILYQEDLSGKQILKASADFLKEIIISTDSNRSLKLIHKEVENALKPPI NTGLE 420
DB 361 LETYHASHILYQEDLSGKQILKASADFLKEIIISTDSNRSLKLIHKEVENALKPPI NTGLE 420
QY 421 RINTRENIQLYNDNTRILKTTYHSSNISNTDYLRILAVEDFYTCQSIYREBELKGLERWV 480
DB 421 RINTRENIQLYNDNTRILKTTYHSSNISNTDYLRILAVEDFYTCQSIYREBELKGLERWV 480
QY 481 ENKLDQKFAKQKTAICYFSVAATLSPELSDARISWAKNGIILTUVDDDFDGGTIDEL 540
DB 481 ENKLDQKFAKQKTAICYFSVAATLSPELSDARISWAKNGIILTUVDDDFDGGTIDEL 540
QY 541 TNLIOQVEKWNVDVDDKCCSEHVRILFLALKDAICWIGDEAFKQWQARDVTSHVITQWLEL 600
DB 541 TNLIOQVEKWNVDVDDKCCSEHVRILFLALKDAICWIGDEAFKQWQARDVTSHVITQWLEL 600
QY 601 MNSMLREAIWTRDAYVPTLNEYMENAYVSPALGPVVKPAIFYVGPGLSEBIVSESYHNL 660
DB 601 MNSMLREAIWTRDAYVPTLNEYMENAYVSPALGPVVKPAIFYVGPGLSEBIVSESYHNL 660
QY 661 FKLMSQGRLLNDIHSFKREFKGLNAVALHLSNGESGKVEBEVVEEMMMIKNKKREL 720
DB 661 FKLMSQGRLLNDIHSFKREFKGLNAVALHLSNGESGKVEBEVVEEMMMIKNKKREL 720
QY 721 MKLIFEENGSI VPRACKDAFWNMCHVNLFFYANDDGTGNTILDTVKDIIYNPLVLNEN 780
DB 721 MKLIFEENGSI VPRACKDAFWNMCHVNLFFYANDDGTGNTILDTVKDIIYNPLVLNEN 780
QY 781 EEQR 784
DB 781 EEQR 784
RESULT 3
US-10-041-018-390
; Sequence 390, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080USI/10025547
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US/10/041.018
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 390
; LENGTH: 788
; TYPE: PRT
; ORGANISM: Lactuca sativa


```

US-10-041-018-395
; Sequence 395, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; PRIOR APPLICATION NUMBER: 2002-01-07
; PRIOR FILING DATE: 2002-01-07
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 395
; LENGTH: 789
; TYPE: PRT
; ORGANISM: Cucurbita maxima
US-10-041-018-395

Query Match          49.2%; Score 2033.5; DB 15; Length 789;
Best Local Similarity 50.0%; Pred. No. 6.2e-142;
Matches 393; Conservative 150; Mismatches 212; Indels 31; Gaps 8;

QY 16 NRPAALSAIHTASTSHGQTNPTNLIIDT-----TKERIQKQKNVEISVSSYD 64
DB 4 SRTGTVARFAASSSSSSASLFGVDVDTTKGALHFPETKERIKLFDKVELSVAYD 63

QY 65 TAWVAVPSPNSPKSPCFPECLNWLNNQNDGSLVNHNNHPL-LKDSLSSTLAC 123
DB 64 TAWVAVPSPNSLNPQLFPFCINWLDSQADSGWLL-----HNQLIMKANLLSTLAC 118

QY 124 IVALKKNVGEDQINKGLSFIEISNLASATEKSQSPIGDIIIPGLLEYAKNLDINLSK 193
DB 119 VLTLRKWNIGHDMSKALDPFKISNLASATDENQSPVGFDIIPFGMIEYAKNLDINLFLA 178

QY 184 QTDPSLMLHRELEOKRCHSNEMDG--YLAIVISEGLNLYDMNMVKYQKMGNSVFNPS 241
DB 179 PTNVDAIVLRKKELELSCRSNSEGKAYLAYVEGIGKLDMDVMQYQREKNSLFPNS 238

QY 242 ATAAAFINHQPCGLNLSLLDKFGNAVTVVPHDLFIRLSMVDITRIERIGISHHPRVEI 301
DB 239 TTAATAFMRNDGCFDRLSLLQKFGDSVPTIYDLYARLHWVDSLOKFGIARHPKEEI 298

QY 302 KNVLDERTYRCWVERDEQIFMDVVTICALAFRLLRINGVEVSPDPLAETINEL-----AL 354
DB 299 RSVLDERTYRCMQGEEINFLDASTCAWAFRLRVEGYDVSSDQITOPSEDIFFNCLGYL 358

QY 355 KDEVAALETYHSHLYQEDLSSGKQILK-SADPLKEIISTD---SNRLSKLIHKEVENA 410
DB 359 KDFGASILEYKASQIITHPDESVLNINSWTSRFLKHGLSDSVSDRTDSVVKQEAUNA 418

QY 411 LKPPINTGLBRINTRNIOQYVNDTRILKTYTHSSNISNTDYLRLAIVEDFYTCQSIYRE 470
DB 419 LEPPYNATLERLISKRAMESYSGDIVRISKSPYACLNFGHQDFLELAIVEDFNTLQRIHLK 478

QY 471 ELKGLERWVENKLDQKFAOKTAYCYFSAVATLSSPELSDARISWAKNGILTTVVDDF 530
DB 479 ELBELQWRWVENKLDLKLFLHILGYCYFAAATLTDPELHDARIAWAQNGVLTVVDDF 538

QY 531 FDIGGTTDELNTLIQCQKKNVVDKDCSEHVRILFLALKDAICWIGDEAFKQWQARDVT 590
DB 539 YDGGSEBELDNLIJELVEKWDPDGEVGYCKDVEIVFLALHSTVCEIGRRALVWQGRSVM 598

QY 591 SHVITQWLELNMGLRBAIWRDAYVPTLNEYMENAVVSFALGHVIVKPAIYFVGPKLSEE 650
DB 599 RNVIDGWLALLKVMRKEAEWSTNKVPSMGTEMQAEHVVSFALGPILLPLFFVGPKLSEE 658

QY 651 IVSSEYHNLFKLMSTQGLINDIHSFKREFKECKLANVALHLNSGSGKVEEBEVEEM 710
DB 659 MIGSEYQKLYKLMSTAGRLKNDIRSDYRECKGKLNILSLMIDGGGNTWKEAIBAIK 718

QY 711 MMTKNRKLMLKLIFFENGSIIVPRACKDAFNNCHVLNFFVANDDDGFTGNLTIDTVKDI 770

```

```

DB 719 GDFERAIRELLGLVLQEN-TTIPRACKOLFVKLMSIVNLFYMEDDGYTSNRLMNTVKAME 777
QY 771 YNPLVL 776
DB 778 EQFMDL 783

RESULT 6
US-10-425-114-65662
; Sequence 65662, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(59313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 65662
; LENGTH: 780
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700265067_FLI.pep
US-10-425-114-65662

Query Match          40.7%; Score 1675; DB 15; Length 780;
Best Local Similarity 44.4%; Pred. No. 6e-116;
Matches 351; Conservative 140; Mismatches 257; Indels 42; Gaps 9;

QY 11 LITKSNRPAALSAIHTASTSHGQTNPTNLIIDTTKERIQKQKNVEISVSSYDTAWVAM 70
DB 1 LLSDSFLPTAYAEERLVAEN---TSLPNMHEKELETIRNQLRKQLPSPSYDTAWVSM 56

QY 71 VPSPNSPKSPCFPECLNWLNNQNDGSLVNHNNHPL-LKDSLSSTLACIVALKRW 130
DB 57 VPRGSHQTPRFPQCVIEWIILNQDDSGWGV-----SQSDSSSVSKDVLSTLACVALKRW 112

QY 131 NVGEDQINKGLSFIEISNLASATEKSQSPIGDIIIPGLLEYAKNLDINLSKQTDPSLM 190
DB 113 NVGRENILWRGLHFIGRNFSVAMDQQTSPIGFNITPGLNLGIDMGLEFPVQTDVCGI 172

QY 191 LHKRELEOKRCHSNEMDG--YLAIVISEGLNLYDMNMVKYQKMGNSVFNPSATAAAF 247
DB 173 LHRREMLKELAVDSFSGKAYNAFIPEGFGNMLDQVMKQKNGSLFSTFTTAVL 232

QY 248 INHQPCGLNLSLLDKFGNAVTVVPHDLFIRLSMVDITRIERIGISHHPRVEIKNVLD 307
DB 233 IHKYNDOALQYLLLYSEFGSAVPAMYPMSVHVQCLSNMVDALERXMGISQRFVSEIESILDM 292

QY 308 TYRCWVERDEQIFMDVVTICALAFRLLRINGVEVSPDPLAETINELAKDEVAALETYHAS 367
DB 293 AYNCWLQKQBEIWMDDATCAWAFRLLRNMGYNVSSDELSHV-----AGASTFSDS 342

QY 368 HILYQEDLSS-----GKOILKSADPLKEIISTDSNRLSK-----LIHKEVEN 409
DB 343 LQGLYNDTKSLLELYKTSKVTLSENDLILDRIGSWSGNLLKDKMCCSRVQKDSIFGMEY 402

QY 410 ALKPPINTGLBRINTRNIOQYVNDTRILKTYTHSSNISNTDYLRLAIVEDFYTCQSIYR 469
DB 403 AVNPFYSTLLERLEHKNIEHFDANGSLML-TTKSSSFRINQBFALAVEDFDFSQSVYR 461

QY 470 EELKGLERWVENKLDQKFAOKTAYCYFSAVATLSSPELSDARISWAKNGILTTVVDD 529
DB 462 DELRHLDSWVENKLDQLOFARQKLYCYLSAAATVPSSELSDARISWAKNGVLTVVDD 521

```

QY 530 FFDIGGTIDELTNLTQVEKKNVDVDDKCCSEHVRILFLALKDAICWIGDEAFKQWADV 589
DB 522 FFDVGSKEELENLALVEKHAHVAVFYSQVIVSAIYTYVNHGAMASAAQGRDL 581
QY 590 TSHVQTWLELWMLRRAIWRDAYVPTLNHMYENAVSPALGPVPAIYFVGPKLSE 649
DB 582 TNHLVEIWLDIRSNMVAEAWQRCQVPTVEYMTNAVVSFALGPVILPALYFVQELLE 641
QY 650 EIVSESHNPLFKLMSTGRLLNDHSKREPKCKLNAVALHLSNGSGKVEEVEEM 709
DB 642 HAVKDEEYDKFLVSTGRLLNDVQSLEREGNCKLMSVLLVHSGSMSIEAAKAM 701
QY 710 MMTMKRKEMLKMLIFEENGSIIVPRACKDAFNMCHVLFNFYANDDGTG-NTIIDLTKD 768
DB 702 QKSIDVSRDLRLVLRKE-SAVPRCKELFWCKILHLFYSQNDGFSSPKEMYSAYNA 760
QY 769 ILYNPLVLVN 778
DB 761 VINBPLKVQN 770

RESULT 7

US-10-259-194A-108
; Sequence 108, Application US/10259194A
; Publication No. US20040010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Ghassemlan, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Krops, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
; CURRENT APPLICATION NUMBER: US/10/259,194A
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,743
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 662
; SOFTWARE: PatentList.pl version 3.0.4 (c) 2001 Syngenta
; SEQ ID NO 108
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-259-194A-108

Query Match 37.4%; Score 1537.5; DB 15; Length 740;
Best Local Similarity 41.7%; Pred. No. 1e-105;
Matches 325; Conservative 142; Mismatches 243; Indels 69; Gaps 16;
QY 20 ALSAIHTASTSHGGQTNPTNLIIDTKERIOKFNVEISVSYDTANVAVPS-PNSPK 78
DB 5 AMRHCSSSSEEGAAATTAARSVRRL--QLAPSPSPYDTANVAVPALRRGG 62
QY 79 SPCPECLNWLNNQDGSW--GLVNHNNHP--LLKSLSTSLACIVALKRWNVGE 134
DB 63 GPRFPQCVAVIQNRQGDGSRHAAAHQQLGSSPEIVTERDLSLSTLACVLARWDAGS 122
QY 135 DQINKGLSFISNLASATEKQSPPIGFDII-FPGLLEYAKMLDNLNLS-----KQTD 187
DB 123 EHVRRGLQFIGRNMVAMDQTAAPASGVSVFAAMLRAMEMGLEVPASQADVRDRDA 182
QY 188 SIMLHKREKQKCHSNEMDGYLAYISEGLNLYDWNVVKYQMKNGSVFNSPSATAAF 247

DB 193 GVI-----CHGGRTY-YTAVSEGLGNIQWNEVMKFORNGSLFNSPYTAAAL 231
QY 248 INHONPGCLNYLNSLLDKFGNAVFPYPHDLFIRLSMVDTIERLGI SHHFVRVEIKNVLD 307
DB 232 VENVDAKALQYLDMLDKFGSAVPAAYPANIQSOLYMDVLEKMGISRFVGEIKSILDM 291
QY 308 TYRCWVEDEOIFMDVVTICALAFRLIRINGVEVSDPDLAEIT-----NELA--LKDEVA 359
DB 292 TYSCKWQKDEEILVDMQTCGAFRLMRNGXDVSDELSHFSEPSFNSLQGYLNDTRS 351
QY 360 ALEYTHASHILYQEDLSSGKQILKSAFLKEITISTDSNRLSKLIHKEVENALKFPINTGL 419
DB 352 LLELHKASKV-----SIAEKEVEYALEFPFYTIL 380
QY 420 BRINTRNIQYVNDVNTIRLKTTHYSSNI SNTDYLRLAVEDFYTCQSYRRELKGLERWV 479
DB 381 DRLDHKRNIHFDTISSQMLETAVLPCH-SNEEIMALGVRDFSSQFIFQBELQQLNSWV 439
QY 480 VENKLDQLKFAOKTAYCYFSVAATLSSPELSDARISWAKNGILTTVVDDDFDIDGTTDE 539
DB 440 KESRLDQLQFARQKLDYFYFSAATITFPELSDVRILWAKNGVLTVVDDDFDVGSGKEE 499
QY 540 LTNLTQVEKKNVDVDDKCCSEHVRILFLALKDAICWIGDEAFKQWADVTSHTVITWLE 599
DB 500 LENLVALVEKWDKNDKTEYVSEQVEIVFSAIYTSNQLGSMASVVGQEDVTKHLVEIQE 559
QY 600 LMNSMLREAIWTRDAYVPTLNHMYENAVSPALGPVPAIYFVGPKLSEIVESEVHN 659
DB 560 LLSRMTVEVWRQSRVYVPTABEYMENAVVTALGPVLPALYLVGPKIPDSVIRSQCSE 619
QY 660 LFKLMSTGRLLNDHSKREPKCKLNAVALHLSNGSGKVE-PEVVEEMMMIKNRK 718
DB 620 LFRMSKGRLLNDVQSYEREGSQGLNSVSL-LALHSGSVSMEEAVKIQRPLEKCR 678
QY 719 ELMKLIFFEENGSIIVPRACKDAFNMCHVLFNFYANDDGTGNTIIDLTKD-IYINPLVL 776
DB 679 ELLKLVVSRGGA-VPRPCRELFWSMCKVCHFFYSGDGFSSPTAKAGALDAVIHEPLNL 736

RESULT 8

US-10-437-963-121564
; Sequence 121564, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kowalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 121564
; LENGTH: 689
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(689)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_24577C.1.pep
US-10-437-963-121564

Query Match 35.4%; Score 1458; DB 16; Length 689;
Best Local Similarity 41.9%; Pred. No. 7.7e-100;

Db 260 FTSQTHQREKQLQKQIPLNVLNHTOALLYSLGLQDVVDWQOEITNLSQSDGSLSSPAS 319
QY 243 TAAAFINONPCGLNVLNLSLLDKFGNAVTVVPHDLFIRLSMVDTIERLIGISHHREVEIK 302
Db 320 TACVFMHTQKRCLEHFNFLSKFGDYVFCYPLDLFEELWAVDVTRIGIDRYFKKEIK 379
QY 303 NVLDETYRCW-VERDE---QIFMDVVTICALAFRLRLRINGEVSPPDLAEITNE----- 351
Db 380 ESLDYVYRWDAERGVGWARCNPDPVDDTAMGLRLRLHGVNVSDDVLENFRDEKGDFF 439
QY 352 -----LALKDEYAALETHASHILYQEDLSSGKOILKSA-----DFLKEIISTD 395
Db 440 CFAQOQIGVTDN---LNLRYCSQVCF-----PGEKIMEEAKTFTTNHLQNALAKKNAPD 491
QY 396 SNRLSLIHKHEVENALKFPINTGLERINTFRNIQLVNDNTRILKTYHSSNISNTDYLR 455
Db 492 KNVKKDLPEGEVEYAIKYPWHRSMPLREARSYIEQFGSNDVWLKTVYKMLYVVSNEKYLE 551
QY 456 LAVEDFVTCQSIYREELKGLERWVENKLDQLKFAKQKTAICYFSVAATLSSPELSDARI 515
Db 552 LAKLDFNMVQALHOKETOHIVSWRESGFNDLFTTQRPVEMYFSVAVSMFEFEFAACRI 611
QY 516 SWAKNGILTFTVVDFFDIDGTIDELNLIQCEKKNVVDKCCSHVRILFLALKDAIC 575
Db 612 AVAKTSLAVILDDLYDTGSLDDLKLFSEAVRRWDISVLDSVRDNQLKVCFTGLYNTVN 671
QY 576 WIGDEAFKQARVTVSHVICTWLELNNMLREAIWTRDAYSPTLNEYMENAYVSEALGPI 635
Db 672 GFCKGLKEQGRDVLGYLRKRWEGLLASYTKEAENSAAKYVPTFNEYVENAKVSIALATV 731
QY 636 VKPAIYFVGPKLSEIEVESSEYHNLF-KLMSTOGRILNDIHSFKRBFKEGKL-NAVALHL 693
Db 732 VLNSIFTGELLDPYILQQVDLRSKFHLVSLTGRLLNDTKYQAEENRGELVSSVQCYM 791
QY 694 -SNGESGKVEEVEEENMMIKRKRKELMKLIFEENGSIIVPRACKDAFWMCHVLNFFYA 752
Db 792 RENPEC--TEEEALSHVYGIIDNALKEL-NWELANPASNAPLCVRRLLENTARVMOLFYM 848
QY 753 NDDGF--TGNITLDYKDIYNPL 774
Db 849 YRDGFGISDKEMKDHVSRTLFDPV 872

Search completed: October 17, 2004, 09:18:45
Job time : 75.5657 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 03:24:39 ; Search time 210.188 Seconds
(without alignments)
1338.060 Million cell updates/sec

Title: US-10-041-018-383
Perfect score: 4113
Sequence: 1 MNLSICIASPLTKSNRPAA.....TVKDIYNPLVNVNEEQR 784

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4113	100.0	784	8	Adm98963 Diterpene
2	4026	97.9	784	8	Adm98964 Diterpene
3	2855.5	69.4	788	8	Adm98970 Diterpene
4	2101.5	51.1	785	8	Adm98969 Diterpene
5	2023.5	49.2	789	8	Adm98975 Diterpene
6	2021	49.1	785	5	Adm98975 Diterpene
7	2012.5	48.9	789	2	Adm98975 Diterpene
8	1537.5	37.4	740	8	Adm98975 Diterpene
9	1523	37.0	732	8	Adm98975 Diterpene
10	1075	26.1	873	6	Adm98975 Diterpene
11	1075	26.1	873	8	Adm98975 Diterpene
12	1062.5	25.8	868	3	Adm98975 Diterpene
13	1062.5	25.8	868	8	Adm98975 Diterpene
14	1062.5	25.8	868	6	Adm98975 Diterpene
15	1051.5	25.6	868	6	Adm98975 Diterpene
16	1015.5	24.7	862	6	Adm98975 Diterpene
17	1013.5	24.6	862	2	Adm98975 Diterpene
18	1013.5	24.6	862	3	Adm98975 Diterpene
19	1013.5	24.6	862	5	Adm98975 Diterpene
20	1013.5	24.6	862	8	Adm98975 Diterpene
21	965	23.5	782	3	Adm98975 Diterpene
22	965	23.5	782	2	Adm98975 Diterpene
23	965	23.5	782	4	Adm98975 Diterpene
24	965	23.5	782	5	Adm98975 Diterpene
25	965	23.5	817	2	Adm98975 Diterpene

26	965	23.5	817	2	AAY06566 Grand fir
27	965	23.5	817	2	AAY06562 Grand fir
28	965	23.5	817	5	Abb79391 Taxadiene
29	965	23.5	817	6	Abu09781 White fir
30	962	23.4	817	2	AAY06568 E-alpha-b
31	925	22.5	458	8	Adm98972 Diterpene
32	863	21.0	870	3	AAY70719 Clarkia b
33	863	21.0	870	6	AAY70719 Clarkia b
34	854.5	20.8	884	5	AAB29780 C. Brewer
35	823	20.0	870	2	Abb91405 Herbicida
36	823	20.0	870	2	Aaw17080 Clarkia b
37	752.5	18.3	811	8	Adm94222 Soybean E
38	743	18.1	815	3	AAY84104 Soybean c
39	743	18.1	815	5	AAY84104 Soybean c
40	732	17.8	787	8	Adm98965 Diterpene
41	725.5	17.6	801	8	Adm98974 Diterpene
42	709	17.2	802	5	AAR90848 Gibberell
43	709	17.2	802	5	Abb92819 Herbicida
44	657.5	16.0	823	8	Adm98976 Diterpene
45	656	15.9	823	2	AAR93800 Anther ea

ALIGNMENTS

RESULT 1

Adm98963
ID ADM98963 standard; protein; 784 AA.

XX AC ADM98963;

XX DT 01-JUL-2004 (first entry)

XX DE Diterpene synthase polypeptide #1.

KW Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;
KW diterpene synthase; defence toxin; volatile defensive signal;
XX pollinator attractant; photoprotectant; enzyme.

XX OS Stevia rebaudiana.

XX PN US2004072323-A1.

XX PD 15-APR-2004.

XX PF 07-JAN-2002; 2002US-00041018.

XX PR 05-JAN-2001; 2001US-0259890P.

XX PA (MATS/) MATSUDA S P T.

XX PA (HART/) HART E A.

XX PI Matsuda SPT, Hart EA;

XX DR WPI; 2004-373921/35.

XX PT New unicellular organisms comprising exogenous nucleic acids encoding a
PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for
producing diterpenes and diterpene precursors.

XX PS Claim 31; SEQ ID NO 383; 38pp; English.

XX CC The invention relates to a unicellular organism for producing a diterpene
or diterpene precursor comprising an exogenous nucleic acid sequence
encoding a geranylgeranyl pyrophosphate synthase under the control of a
promoter operable in the organism, and an exogenous nucleic acid sequence
encoding a diterpene synthase under the control of a promoter operable in
the organism. The invention also relates to methods of producing a
diterpene or diterpene precursor and a method of isolating a diterpene
synthase comprising growing several cells in the presence of a
polyaromatic resin to make a cell/resin mixture, where at least one of
the cells further comprises at least one isolated and purified nucleic
acid sequence of a yeast expression library, and the expression of the

CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents a diterpene synthase polypeptide of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html.

XX	Sequence 784 AA;	
SQ	Query Match	100.0%; Score 4113; DB 8; Length 784;
	Best Local Similarity	100.0%; Pred. No. 0;
	Matches 784; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1	MNLSICTASPLLTCKSNRPAALSAIHTASTSHGGQTNTNLIIIDTTKRIQKQFNVEISV 60
DB	1	MNLSICTASPLLTCKSNRPAALSAIHTASTSHGGQTNTNLIIIDTTKRIQKQFNVEISV 60
QY	61	SSYDTAWAVSPSPKSPCEPCCLNLLNQLNDGSLVNHHTNHNHPLIKDLSST 120
DB	61	SSYDTAWAVSPSPKSPCEPCCLNLLNQLNDGSLVNHHTNHNHPLIKDLSST 120
QY	121	LACIVALKRWNVGEDQINKGLSPFIESNLASATEKSPSPIGFDIIIPFGLLEYAKNLDINL 180
DB	121	LACIVALKRWNVGEDQINKGLSPFIESNLASATEKSPSPIGFDIIIPFGLLEYAKNLDINL 180
QY	181	LSKQTFSLMLHKKRELEQKRCHEMDGYLAIYISEGLGNLYDNWVKKYQMKNGSVFNSP 240
DB	181	LSKQTFSLMLHKKRELEQKRCHEMDGYLAIYISEGLGNLYDNWVKKYQMKNGSVFNSP 240
QY	241	SATAAFINHQPCCLNLYSLDKFCGNVPTVPHDLFRLSWDTIERLGISSHFRVE 300
DB	241	SATAAFINHQPCCLNLYSLDKFCGNVPTVPHDLFRLSWDTIERLGISSHFRVE 300
QY	301	IKNVLDETVCWVERDEQIFMDVVTTCALAFRLRINGEYVSPDPLAIBITNELAKDEYAA 360
DB	301	IKNVLDETVCWVERDEQIFMDVVTTCALAFRLRINGEYVSPDPLAIBITNELAKDEYAA 360
QY	361	LETYHSHILYQEDLSGKQILKSADFLKEIISTDSNRLSKLIHKEVENALKFPINTGLE 420
DB	361	LETYHSHILYQEDLSGKQILKSADFLKEIISTDSNRLSKLIHKEVENALKFPINTGLE 420
QY	421	RINTRNLIQYVNDTRILKTTVHSSNISNTDYLRLAIVEDFYTCQSIYREELKGLRWVV 480
DB	421	RINTRNLIQYVNDTRILKTTVHSSNISNTDYLRLAIVEDFYTCQSIYREELKGLRWVV 480
QY	481	ENKLDQKFAKQRTAYCYFVAATLSSPELSDARISWAKNGILTTVVDDFDIGGTIDEL 540
DB	481	ENKLDQKFAKQRTAYCYFVAATLSSPELSDARISWAKNGILTTVVDDFDIGGTIDEL 540
QY	541	TNLIQCKEKNVDKDCSEHVRILPLAKDAICWIGDAFQWQARDVTSVHVIQWLLEL 600
DB	541	TNLIQCKEKNVDKDCSEHVRILPLAKDAICWIGDAFQWQARDVTSVHVIQWLLEL 600
QY	601	MNSMLREAIWTRDAYVPTLNEYMENAVSPALGPVFKPAIFVGPKLSEBIVSESYHNL 660
DB	601	MNSMLREAIWTRDAYVPTLNEYMENAVSPALGPVFKPAIFVGPKLSEBIVSESYHNL 660
QY	661	FKLMSTQGRLLNDIHSFKRBFKEGKLNALVALHLSNGSGKVEEBVVEEMMMINKKREKEL 720
DB	661	FKLMSTQGRLLNDIHSFKRBFKEGKLNALVALHLSNGSGKVEEBVVEEMMMINKKREKEL 720
QY	721	MKLIIFENGSTIVPRACKDAFNMWCHVLNFFVANDDGTGTNTLTDVXDIYINPLVLNEN 780
DB	721	MKLIIFENGSTIVPRACKDAFNMWCHVLNFFVANDDGTGTNTLTDVXDIYINPLVLNEN 780
QY	781	EEQR 784

DB	781	EEQR 784
RESULT 2		
ADM98964		
ID	ADM98964	standard; protein; 784 AA.
XX	ADM98964;	
AC	ADM98964;	
XX	01-JUL-2004	(first entry)
DT		
XX	Diterpene synthase polypeptide #2.	
DE		
XX	Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;	
KW	diterpene synthase; defence toxin; volatile defensive signal;	
KW	pollinator attractant; photoprotectant; enzyme.	
XX	Stevia rebaudiana.	
XX	US2004072323-A1.	
PN		
XX	15-APR-2004.	
PD		
XX	07-JAN-2002; 2002US-00041018.	
XX	05-JAN-2001; 2001US-0259880P.	
XX	(MATS/) MATSUDA S P T.	
PA	(HART/) HART E A.	
XX	Matsuda SPT, Hart EA;	
PI		
XX	WPI; 2004-373921/35.	
DR		
XX	New unicellular organisms comprising exogenous nucleic acids encoding a	
PT	geranylgeranyl pyrophosphate and a diterpene synthase, useful for	
PT	producing diterpenes and diterpene precursors.	
XX		
XX	Claim 31; SEQ ID NO 384; 38pp; English.	

XX The invention relates to a unicellular organism for producing a diterpene
CC or diterpene precursor comprising an exogenous nucleic acid sequence
CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene
CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the
CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents a diterpene synthase polypeptide of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html.

XX	Sequence 784 AA;	
SQ	Query Match	97.9%; Score 4026; DB 8; Length 784;
	Best Local Similarity	98.0%; Pred. No. 0;
	Matches 768; Conservative	8; Mismatches 8; Indels 0; Gaps 0;
QY	1	MNLSICTASPLLTCKSNRPAALSAIHTASTSHGGQTNTNLIIIDTTKRIQKQFNVEISV 60

Db 1 MNLSLCIASPLLTAKSRPTALSAIHTASTSHGGQNTPTNLIIDTTKERIQKLFKNVEISV 60
Qy 61 SSYDTAWAVWVSPNSPKSPCEPCINLWLNQNDGSLVNNHNNHHPKLLKDSLSST 120
Db 61 SSYDTAWAVWVSPNSPKSPCEPCINLWLNQNDGSLVNNHNNHHPKLLKDSLSST 120
Qy 121 LACIVALKENWVGEQINKGLSFIESNLASATEKSPSPGPDIIIPGGLLEKAKNLDNL 180
Db 121 LACIVALKENWVGEQINKGLSFIESNLASATEKSPSPGPDIIIPGGLLEKAKNLDNL 180
Qy 181 LSKQTDLSMLHRELEOKRCHSNEMDGYLAVISEGLGNLYDNWVKKYQMKNGSVFNSP 240
Db 181 LSKQTDLSMLHRELEOKRCHSNEMDGYLAVISEGLGNLYDNWVKKYQMKNGSVFNSP 240
Qy 241 SATAAFTNHQNGPCINLYNSLLDFGNAPVVPYPHDLFIRLSMVDTTIRLGI SHHFRVE 300
Db 241 SATAAFTNHQNGPCINLYNSLLDFGNAPVVPYPHDLFIRLSMVDTTIRLGI SHHFRVE 300
Qy 301 IKNVLDITYRQVVERDEQIFMDVVTICALAFRLRLRINGEVSDDPLAETINELALKDEYAA 360
Db 301 IKNVLDITYRQVVERDEQIFMDVVTICALAFRLRLRINGEVSDDPLAETINELALKDEYAA 360
Qy 361 LETHASHILOEDLSSGKQILKSDFLKBIISTDSNRLSKLIHKEVENALKFPINTGLE 420
Db 361 LETHASHILOEDLSSGKQILKSDFLKBIISTDSNRLSKLIHKEVENALKFPINTGLE 420
Qy 421 RINTFRNTQLYNVNDTRILKTYHSSNINTDYRLAIVEDFTTCSYIREELKGLERWV 480
Db 421 RINTFRNTQLYNVNDTRILKTYHSSNINTDYRLAIVEDFTTCSYIREELKGLERWV 480
Qy 481 ENKLDQLFARQKAYCYFVSVAATLSSPELSDARISWAKNGILTTVDVDFDGTIDEL 540
Db 481 ENKLDQLFARQKAYCYFVSVAATLSSPELSDARISWAKNGILTTVDVDFDGTIDEL 540
Qy 541 TNLQCVKKNVVDKDCSEHVRILFLALKDAICWIGDEAFKQARDVTSHVITWLEL 600
Db 541 TNLQCVKKNVVDKDCSEHVRILFLALKDAICWIGDEAFKQARDVTSHVITWLEL 600
Qy 601 MNSMLREAITWDRADVPTLNYMENAYVSPALGPVKAIFYVGPKLSEEVSEYHNL 660
Db 601 MNSMLREAITWDRADVPTLNYMENAYVSPALGPVKAIFYVGPKLSEEVSEYHNL 660
Qy 661 FKLMSTOGLRLNDIHSFKREPKKGLNAVALHLSNGESGKVEEVEVEMWMIKNKREL 720
Db 661 FKLMSTOGLRLNDIHSFKREPKKGLNAVALHLSNGESGKVEEVEVEMWMIKNKREL 720
Qy 721 MKLIFEENGSIIVPRACKDAFNMCHVLMFFYANDDGTNTILDTVKDIIYNPLVNNEN 780
Db 721 MKLIFEENGSIIVPRACKDAFNMCHVLMFFYANDDGTNTILDTVKDIIYNPLVNNEN 780
Qy 781 EEQR 784
Db 781 EEQR 784

RESULT 3

ADM98970

ID ADM98970 standard; protein; 788 AA.

AC

XX

XX

DT 01-JUL-2004 (first entry)

DE Diterpene synthase polypeptide #8.

KW Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;
KW diterpene synthase; defence toxin; volatile defensive signal;
KW pollinator attractant; photoprotectant; enzyme.

OS Lactuca sativa.
XX US2004072323-A1.

XX 15-APR-2004.
PD
XX 07-JAN-2002; 2002US-00041018.
PF
XX 05-JAN-2001; 2001US-0259880P.
PR
XX (MATS/) MATSUDA S P T.
PA (HART/) HART E A.
FA
XX Matsuda SPT, Hart EA;
PI WPI; 2004-373921/35.
XX
DR
XX New unicellular organisms comprising exogenous nucleic acids encoding a
PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for
PT producing diterpenes and diterpene precursors.
XX
PS Claim 31; SEQ ID NO 390; 38pp; English.
XX
CC The invention relates to a unicellular organism for producing a diterpene
CC or diterpene precursor comprising an exogenous nucleic acid sequence
CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene
CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the
CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents a diterpene synthase polypeptide of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 788 AA;

Query Match

Best Local Similarity 69.4%; Score 2855.5; DB 8; Length 788;

Matches 541; Conservative 102; Mismatches 116; Indels 39; Gaps 6;

Qy 1 MNLSLCIASPLLTAKSN-----RPAALSIAHTASTSHGGQNTPTNLIIDT 44
Db 1 MNIAQTSSAMLVPSHSI PHRSWVWVNCWQYNP---SGLRTAS-SQAGQVNPVTWLDV 56
Qy 45 TKERIQKLFKNVEISVSSYDTAWAVWVSPNSPKSPCEPCINLWLNQNDGSLVNNH 104
Db 57 TKERIRKLFNNVEISVSSYDTAWAVWVSPNSPKSPCEPCINLWLNQNDGSLVNNH 114
Qy 105 THNNHPLKDSLSSTLACTIVALKRWNVGDDQNKGLSFIESNLASATEKSPSPGPD 164
Db 115 --PHQSPLIKDTLSSTLACTIVALKRWNVGDDQNKGLSFIESNLASATEKSPSPGPD 172
Qy 165 IFPGLEAYAKNLDJNLKSKQTDLSMLHRELEOKRCHSNEMDGYLAVISEGLGNLYDN 224
Db 173 IFPGLEAYAKNLDJNLKSKQTDLSMLHRELEOKRCHSNEMDGYLAVISEGLGNLYDN 232
Qy 225 MVKQYQMKNGSVFNSPSPATAAFTNHQNGPCINLYNSLLDFGNAPVVPYPHDLFIRLSM 284
Db 233 MVKQYQMKNGSVFNSPSPATAAFTNHQNGPCINLYNSLLDFGNAPVVPYPHDLFIRLSM 292
Qy 285 VDTIERLGI SHHFRVEIKNVLDITYRQVVERDEQIFMDVVTICALAFRLRLRINGEVS 344

Db 293 VDTLRLGIGKRFHMFVEIQNVLDYRCVWQGVQIFMDVVTICALAFRLVRSNGYVSSDP 352
 QY 345 LAETN-----FLAKDEYAALETYHSHILYOEELSSGKOLKSGADFLKEIISTDS 396
 Db 353 LAKITKEGYMNSPEKPKDVTSLEVYKASQIIYOEELAFREQLTS-----YLPSS 405
 QY 397 NLSKLIHKEVENALKFPINTGLERINTRRNIQLYNVNTRILKTYHSSNISNTDYLR 456
 Db 406 NKLSNLIKVEDDALKFPNGSLRMSFRNIEHYNLHNEHTRILKTYSSNISNKLK 465
 QY 457 AVEDFYTCQSIYREELKGLRWVVENKLDQLKPARQKTAICYPSVAATLSSPELSDARIS 516
 Db 466 AVODFNEQOSIYCEELKOLRWVVENRLDKLKFARQKTAICYPSAATLSSPDLSARIS 525
 QY 517 WAKNGILTVVDDPDIIGTIDELTNLIQCVEKNVVDKCCSEHVRILFLAKDAICW 576
 Db 526 WAKSSILTVVDDPFDVGGSDDELNVFHHIIBKNVNVVENDCCSEVGVFLAKDAICW 585
 QY 577 IGDEAFKQWQARDVTSVHVIQTLWELMNSMLREAIWTRDAYVPTLNEYMENAYVSFALGPV 636
 Db 586 IGDKAFKIQERNITSHVIEIWDVKSMLREALWAKDGSIPTINEYMWENGYVSFALGPV 645
 QY 637 KPAIVFVGKLSSEIVSESEYHNLFKLMSQORLNDIHSFKREPKGKLNALVALHLSNG 696
 Db 646 LPTLYFLGVKLSSEVVQSSEYHKLVEVMSQORLNDIHSFKREKAGKLNALVALYMSDG 705
 QY 697 ESKGVEEVEEVMWMMIKNKKELMKLIPEENGSIYPRACKDAFNMCHVLANFEYANDDG 756
 Db 706 KSGSEVEEVEEMKILTSQRKEMMKLVLETKGVSVPVCKOVFNMCMVNLNLFYATDDG 765
 QY 757 FTGNITLDVKDIYNPL 774
 Db 766 FTGNALDVVKELIYEPV 783

RESULT 4
 ADM98969
 ID ADM98969 standard; protein; 785 AA.
 AC ADM98969;
 AC ADM98969;
 XX
 DT 01-JUL-2004 (first entry)
 XX
 DE Diterpene synthase polypeptide #7.
 XX
 KW Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;
 KW diterpene synthase; defence toxin; volatile defensive signal;
 KW pollinator attractant; photoprotectant; enzyme.
 XX
 OS Cucumis sativus.
 XX
 PN US2004072323-A1.
 XX
 PD 15-APR-2004.
 XX
 PF 07-JAN-2002; 2002US-00041018.
 XX
 PR 05-JAN-2001; 2001US-0259880P.
 XX
 PA (MATS/) MATSUDA S P T.
 PA (HART/) HART E A.
 XX
 XX Matsuda SPT, Hart EA;
 XX
 DR WPI; 2004-373921/35.
 XX
 XX New unicellular organisms comprising exogenous nucleic acids encoding a
 PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for
 PT producing diterpenes and diterpene precursors.
 XX
 PS Claim 31; SEQ ID NO 389; 38pp; English.
 XX
 CC The invention relates to a unicellular organism for producing a diterpene

CC or diterpene precursor comprising an exogenous nucleic acid sequence
 CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
 CC promoter operable in the organism, and an exogenous nucleic acid sequence
 CC encoding a diterpene synthase under the control of a promoter operable in
 CC the organism. The invention also relates to methods of producing a
 CC diterpene or diterpene precursor and a method of isolating a diterpene
 CC synthase comprising growing several cells in the presence of a
 CC polyaromatic resin to make a cell/resin mixture, where at least one of
 CC the cells further comprises at least one isolated and purified nucleic
 CC acid sequence of a yeast expression library, and the expression of the
 CC nucleic acid sequence is regulated by an inducible promoter under
 CC conditions where the expression is induced, filtering the cell/resin
 CC mixture, extracting the cell/resin mixture with alcohol to produce an
 CC organic eluent and analysing the organic eluent by a screening method
 CC including chromatography and/or spectroscopy, to identify the nucleic
 CC acid sequence encoding the diterpene synthase. The unicellular
 CC microorganism is useful as a diterpene or diterpene precursor producing
 CC system. Diterpenes, in plants, serve as defence toxins, volatile
 CC defensive signals, pollinator attractants and photoprotectants. This
 CC sequence represents a diterpene synthase polypeptide of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format from USPTO at
 CC seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 785 AA;
 Query Match 51.1%; Score 2101.5; DB 8; Length 785;
 Best Local Similarity 50.9%; Pred. No. 1.7e-162;
 Matches 413; Conservative 136; Mismatches 194; Indels 69; Gaps 10;
 QY 1 MNLSLCIASPLTKSNRPAALSAIHTASTSH-----GGQTNFNLIIDTKERIQK 52
 Db 1 MNLS-----RPTNLGCF-TASSASLFFPLDVGTTKTGALFEETKEIKKL 47
 QY 53 FKNVETSVSYDTAVWVVPSPNSPKSPCPFCPLNLIINNOLDGSGWLNTHNHPL 112
 Db 48 FKNVETSVSYDTAVWVVPSPNSLNLKPLFPCCINWLDHQNDPSGWGL-----HDHQL 102
 QY 113 -LKDSLSTLACTIVALKRWNVGEDIQKGLSFIESNLASATEKSPSPGDIIFPGLLE 171
 Db 103 VMKATLLSLTACVLTJLKRWDIGDDHMSKALSFIKSNIASATDENQSPVDFIIFPMIE 162
 QY 172 YAKNLDINLLSKQTFSLMLHRELEQKCHENMDG--YLAIVSEGLNLDWNNVKKY 229
 Db 163 YAKNLDINLLPLASMNVDALVQKLELRSCSNSEGGKAYLAYVSEGLQDQWENVMRY 222
 QY 230 QMKGVSFNSPSATAAFINHQNPCLNLSLDKFGNAVPTVYPHDLFIRLSMVDTE 289
 Db 223 QRRNGSLFSSPSTTAVAFMRNRDDGCFNVRSLQKPHSVPAIYPLDIYARLHWVDSLQ 282
 QY 290 RLGISHFVEIKNVLDETYRCWVERDEQIFMDVVVTCALAFRLLRINGEVSPDPAEIT 349
 Db 283 KLIGIDGHFKDEIRSVLDETYSCWMQGEENIFLDASTCAMAFLRVEGYDVSSDQLTQFS 342
 QY 350 NEL-----ALKDEYAALETYHSHILYOE-----LSSGKQILKS 384
 Db 343 EGLFNSCLGHLKDFGSALELFAKSOIITYPDEFILNINSWTSRFINHLSGSG----- 396
 QY 385 ADPLKEIITDSNRLSKLIHKEVENALKKPPINTGLERINTRRNIQLYNVNTRILKTYH 444
 Db 397 -----SVHSDRTERLVKQEAUNAFEPFYNSTLERLSNKRALYESYGDIVLSKTAYA 448
 QY 445 SSNISMNTDYRLAVEDFYTCQSIYREELKGLRWVVENKLDQLKPARQKTAICYPSVAAT 504
 Db 449 CLNFGHQDFLELAVEDFNTLQIHRKELKELEKWTENKLDKLKPARQKTAICYPSAAAT 508
 QY 505 LSSPELSDARISWAKNGILTVVDDPFDIGTIDELTNLIQCVEKNVVDKCCSEHVR 564
 Db 509 LTSPELSDARLSWAKNGVLTTVVDDPFDVGGSEELVNLIIQLVERKWDASGETGYCSKEVE 568
 QY 565 ILFLALKDAICWIGDEAFKQWQARDVTSVHVIQTLWELMNSMLREAIWTRDAYVPTLNEYME 624
 Db 569 IIFLALHSTICEIGKALKALPWQGSVNRNVIDIWLALLESNRKEAEMLKNKVVSFLDEYME 628

XX Herbicidally active polypeptide SEQ ID NO 874.
XX DE
XX KW Herbicidal; plant; agriculture; herbicide.
XX OS Arabidopsis thaliana.
XX PN WO200210210-A2.
XX PD 07-FEB-2002.
XX 28-AUG-2001; 2001WO-EP009892.
XX 28-AUG-2001; 2001WO-EP009892.
XX PA (PARB) BAYER AG.
XX Tietjen K, Weidler M;
XX WPI; 2002-269010/31.
XX Identifying plant target proteins for herbicidally active compounds.
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms.
XX Claim 5; SEQ ID NO 874; 261pp + Sequence Listing; English.
XX The invention relates to identifying target proteins (ABB90790-ABB94016)
CC for herbicidally active compounds, comprising aligning and comparing
CC nucleic acid or amino acid sequences from plant with nucleic acid or
CC amino acid sequences from non-plant organisms using suitable search
CC parameters, where plant sequences having an E-value greater by a factor
CC of 3 than the E-value of most similar non-plant sequences are selected.
CC The polypeptides or nucleic acids encoding them are useful for
CC identifying modulators. The identified modulators are useful as
CC herbicides
XX Sequence 785 AA;
SQ
Query Match 49.1%; Score 2021; DB 5; Length 785;
Best Local Similarity 52.6%; Pred. No. 6.6e-156;
Matches 402; Conservative 130; Mismatches 208; Indels 24; Gaps 9;
QY 34 QTNPTLIIDTTKRIQKQFNVEISVSYDTAWAVPSPNSPKPCFPECLNLIINQ 93
DB 27 QTRANNVFEQTEKIRKMLEKVELSVAYDSWAVPSPSQAPLPFQCWLLDNQ 86
QY 94 LNDGSGLVNHTNHNHPLKDSLSSTLACTIVALKEMNVGEDIKGLSFIESNLASAT 153
DB 87 HEDGSGWLDNHDH--QSLKDVLSSTLASILALKKWGIGERQINKGLQFIELNSALVTD 143
QY 154 KQSPSPIGDITFPGLLEVAK--NLIDNLSKOTDFSLMLHKEELQKCHSNEM----D 207
DB 144 ETIQKTFGTDFIIPFGIKYARDNLNLTPIGSEVVDD--MIRKDLDLK-CDSEKFKSGRE 200
QY 208 GYLAVISEGLNLYDNWVKYQKNGSVSPNSATAAAPHNQPCCLNLYNLSLDKFG 267
DB 201 AYLAIVLEGTRNLKWDNLIVKYQKNGSLDFSPATTAATQFGNDGCLRYLCSLQKPE 260
QY 268 NAVPTVYPHDLFRLSMVDITIELGISHHPRVEIKVLDITYPCWVERDEQIFMDVVTCA 327
DB 261 AAVPSVYPPDQVRLSIIVTLESIGDRDPKTEIKSILDTYRYLWLGDBEICLDLATCA 320
QY 328 LAFRLIRINGEYSPDLAHTWELA-----LXDEVAALTYHSHLYQEDLSGK 379
DB 321 LAFRLILAHGYDVSVDLPKPFABESGFSDTLGVVKNFTSVLELFAAQ-SYPHESALKK 379
QY 380 QILKSADFLEKIIIST--DSNRLSKLIHKEVENALKFPINTGLERINTRRNIQYN-VDNT 436
DB 380 QCCWTKQYLEMEISSLWSVKTSVRDKYLKKEVEDALAFPSYASLERSDHRKILINGSVENT 439
QY 437 RILKTYHSSNISNTDYLRLAIVEDFTYCQISYREELKGLERWVVENKLDQLKFAKQTAY 496

DB 440 RVTKTSYRLHNICTSDILKLAVDDEFQCSIHREEMERLDRWIVENRLOELKFAQKLAY 499
QY 497 CYFESAATLSPPELSDARISWAKNGILTTVVDDFDIGDTIDELNLIQCCKMNVVDK 556
DB 500 CYFESAATLSPPELSDARISWAKGVLTVVDDFDVGGSKKELENLHLVEKNDLNGVP 559
QY 557 DCCSEHVRILFLAKDAICWIGDEAFKWQARDVTSHVIQTWLELMNSMLREAIWTRDAY 616
DB 560 EYSSEHVEIIFSVLRDTILETGKAFYQGRNVTHIVKWLKLLKSMLEAEWSDDKST 619
QY 617 PTINYMENAYVSPALGPVPAIVFVGPKLSEBIESESEYHNLKLMSTQGRLLNDIHS 676
DB 620 PSLEDYMENAYISFALGPVLPATYILIGPELPEKTVDSHQYNQLYKLVSTWGRLLNDIQG 679
QY 677 FKREFKEGKINAVLHLSNGESGKVEEVEEMMMKMKRKLMLKLIPEENGSIIVPRAC 736
DB 680 FKRESAEGKINAVSLHMKHERDNRKSEVIESKGLAERKRELHKLVLLEKGSVVPREC 739
QY 737 KDAPWMMCHLVNFFYANDDGTGNTILDTVKDIIYNPLVLVNNEN 780
DB 740 KEAFLKMSKVLNLFYRKDDGFTSNDMLSLKSVIYEPVSLQES 783
RESULT 7
AAW14055
ID AAW14055 standard; protein; 789 AA.
XX AC AAW14055;
XX DT 17-OCT-2003 (revised)
XX DT 27-JUN-1997 (first entry)
XX DE Pumpkin ent-kaurene B synthase.
XX Ent-kaurene B synthase; plant growth; seed growth; transgenic plant;
KW pumpkin.
XX Cucurbita maxima; L cv. Riesenmelone gelb vernetzt.
XX EP768381-A2.
XX PD 16-APR-1997.
XX PF 07-OCT-1996; 96EP-00116005.
XX PR 09-OCT-1995; 95JP-00261147.
XX PA (RIKA) INST PHYSICAL & CHEM RES.
XX PI Kamiya Y, Yamaguchi S;
XX WPI; 1997-214817/20.
XX N-PSDB; AAT60339.
XX Pumpkin ent-kaurene B synthase polypeptide - used for the regulation of
PT plant and seed growth.
XX Claim 1; Page 13-19; 24pp; English.
XX Pumpkin ent-kaurene B synthase (AAW14055) catalyses the conversion of
CC copalyl pyrophosphate to ent-kaurene and has important functions in the
CC growth of plants and seeds. The enzyme has been purified from pumpkin
CC endosperm, and primers based on isolated peptides (see also (see also
CC AAW14056-60) were utilised in the isolation of an ent-kaurene B synthase
CC cDNA clone (AAT60339). The enzyme can be produced in transformed host
CC cells, and DNA encoding the enzyme can be used to produce transgenic
CC plants of increased height. (Updated on 17-OCT-2003 to standardise OS
XX field)
XX Sequence 789 AA;
SQ
Query Match 48.9%; Score 2012.5; DB 2; Length 789;

XX	Sequence 740 AA;	
SQ	Query Match	
	Best Local Similarity	37.4%; Score 1537.5; DB 8; Length 740;
	Matches 325; Conservative 142; Mismatches 243; Indels 69; Gaps 16;	
QY	20 ALSAHTASTSGGQTNPNTLIDITKRIQKFNVEISVSSTAWVMPSPNSPK 78	
DB	5 AMRHSSSSSEGGAAATAARSVREKL--QLAPSPSPPYDTAWVAMPALRRGG 62	
QY	79 SPFCPECLNWLNNQLDGSW--GLVNHHTNHP--LLKDSLSSTLACIVALKRNVGE 134	
DB	63 GRFPQCVAWIQNRQDGSWRHAAAHQQLGSSPEIVTERDLSSTLACVLARWDAGS 122	
QY	135 DQNKGLSTIENLASATKSPSPGPDII--PPGLLEYAKMLDINLLS-----KQTFP 187	
DB	123 EHVRRGLQFIGRNSVAMDQTAAPASGVSFPAAMLRMAMEMGLEVPASVQADVRDRA 182	
QY	188 SLMLHKRELQKRECHSNEMDGYLAVISEGLNLYDNMVKYQMKNGSVFNPSPATAAF 247	
DB	183 GVI-----CHGRTE-YTAVYSEGLNIQNNWVKFQKNGSLFNSPYTTAAL 231	
QY	248 INHONGCLNLYNSLLDKFNNAVPTVYPHDLFIRLSMVDTIERLGHSHFRVEIKNVLD 307	
DB	232 VEHYDAKALQYLDMLDKGSAVPAAYPANIQSQYVWVDLEKMGISRHFVGEIKSILDM 291	
QY	308 TTRCWVERDQIFMDVVTCALAFRLIRINGEYVSPPLAEIT-----NELA--LKDEYA 359	
DB	292 TYSCKQRDEEIVLDQTCGMAFRLRMNGYDVSSELSHFPSEPSFHNLSQGLYNDTKS 351	
QY	360 ALETHASHILYQEDLSGGQILKSADFLKEIISTDSNRLSKLIHKEVENALKFPNTGL 419	
DB	352 LLELHKASKV-----STAKEVEYALFFPYTIL 380	
QY	420 ERINTRNQLNVNDTRILKTHYSNTSDTYLRLAVEDFYTCQSIYREELKGLRWV 479	
DB	381 DRLDHKNRIEHPDITSSQMLETAYLPCH-SNEEIMALGVDRDFSSQFIQEEELQQLNSWV 439	
QY	480 VENKLDQFARQKTAFCVSVAAATSSPBLSDARISWAKNGILTTVDDPFDIGTIDE 539	
DB	440 KESRLDQFARQKLDYFYFSAATITFPELSDVRILNAKNGVLLTVDVDFDVGSKKE 499	
QY	540 LTWLIQVCKWNVVDKDCSEHVRILFLALDKAIWIGDEAPKQARDVTSHVITWLE 599	
DB	500 LEMVALVEKWDKNDKTEYSEQVEIVFSAYITSTQLGSMASVVGQDVTXHLVIWQE 559	
QY	600 LMNSMLREAIWTRDAYPTLNEYMENAYVSFALGPVVKPAIYVGPVKLSEIYESEYHN 659	
DB	560 LLRSNMTEVEWRQSRVPTAEYVEMNAVVTFALGPVLPALYLVGPKIPDSVIRSQECSE 619	
QY	660 LFKLMSTOGLLNDIHSFKREPEKGNALVALHLSNGESQKVB-EEVVEEMMMIKNRK 718	
DB	620 LFLMELSKGGLLNDVQSYEREGSQGLNSVSL-LALHSGGSVSNEEAVKQIQPIEKRR 678	
QY	719 ELMKLIFEENGSTVPRACKDAFNMCHLVNFFYANDDGFNTILDTVKD-IYINPLVL 776	
DB	679 ELLKLIVSRGGA-VPKPRCLFWSMKVCHFFISGGDGFSSPTAKAGALDAVIEPLNL 736	

RESULT 9
ADM94224
ID ADM94224 standard; protein; 732 AA.
XX
AC ADM94224;
XX
DT 17-JUN-2004 (first entry)
DE Corn Ent-Kaurene Synthase B protein #1.
XX
KW Corn; plant; plant metabolism;
KW GTP cyclohydrolase II/3,4-dihydroxy-2-butanone-4-phosphate synthase;
KW Dioxxygenase; Ent-Kaurene Synthase A; GA-20 oxidase;

KW Gibberellin 3-beta hydroxylase; Riboflavin specific deaminase;
KW Ethylene response factor; Acyl-CoA thioesterase II;
KW ABC transporter GCN20-like; P-glycoprotein I; P-glycoprotein 3;
KW P-glycoprotein ATPGP; P-glycoprotein HVMR2; ABC transporter;
XX PMP70 ABC transporter; MRP4 ABC transporter; transgenic.
XX Zea mays.
OS US6677502-B1.
PN 13-JAN-2004.
PD 12-JUL-2000; 2000US-00614912.
XX 12-JUL-1999; 99US-0143401P.
XX 12-JUL-1999; 99US-0143412P.
PR 30-JUL-1999; 99US-014650P.
PR 15-DEC-1999; 99US-0170906P.
PR 21-DEC-1999; 99US-0172946P.
PR 21-DEC-1999; 99US-0172959P.
XX (DUPO) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX Allen SM, Kinney AJ, Rafalski JA, Orozco EM, Miao G, Famodu OO;
PI Lee J, Lohman KN, Rendina AR, Sakai H, Weng Z, Caimi PG, Fang Y;
PI Shen JB, Zoughi IL, Anderson SL, Shi J, Lu G, Helentjaris TG;
PI Li CP;
DR WPI; 2004-088430/09.
XX N-PSDB; ADM94223.
DR New isolated GTP cyclohydrolase II/3,4-dihydroxy-2-butanone-4-phosphate
PT synthase nucleic acid and proteins, useful for creating transgenic plants
PT where polypeptides are present at higher or lower levels.
XX Example 3; SEQ ID NO 24; 186pp; English.
XX The invention relates to an isolated polynucleotide comprising a
CC nucleotide sequence encoding a polypeptide having GTP cyclohydrolase
CC II/3,4-dihydroxy-2-butanone-4-phosphate synthase activity, appearing as
CC ADM94266. Also included are a vector comprising the polynucleotide, a
CC recombinant DNA construct comprising the polynucleotide operably linked
CC to at least one regulatory sequence, transforming a cell with the
CC polynucleotide, a cell comprising the recombinant DNA construct,
CC producing a plant (comprising transforming a plant cell with the
CC polynucleotide and regenerating a plant from the transformed plant cell),
CC a plant comprising the recombinant DNA construct and a seed comprising
CC the recombinant DNA construct. Also disclosed as new are nucleic acid
CC (cDNA) fragments and assembled contigs encoding plant metabolism proteins
CC chosen from GTP cyclohydrolase II/3,4-dihydroxy-2-butanone-4-phosphate
CC synthase, Dioxxygenase, Ent-Kaurene Synthase A, GA-20 oxidase, Gibberellin
CC 3-beta hydroxylase, Riboflavin specific deaminase, Ethylene response
CC factor, Acyl-CoA thioesterase II, ABC transporter, GCN20-like, P-
CC glycoprotein I, P-glycoprotein 3, P-glycoprotein ATPGP, P-glycoprotein
CC HVMR2, ABC transporter, PMP70 ABC transporter and MRP4 ABC transporter.
CC The nucleic acid fragments may be used to create transgenic plants where
CC the polypeptides are present at higher or lower levels than normal or in
CC cell types or in developmental stages in which they are not normally
CC found. The polynucleotides can be used as probes for genetically and
CC physically mapping genes and as markers for traits linked to those genes.
CC The nucleic acid fragments may be used as restriction fragment length
CC polymorphism (RFLP). It can also be used to probe Southern blots
CC containing restriction endonuclease-treated genomic DNAs of a set of
CC individuals representing parent and progeny of a defined genetic cross.
CC The nucleic acid fragments may also be used as hybridisation probes
CC against PCR amplification products generated from the mutation population
CC using the mutation tag sequence primer in conjunction with an arbitrary
CC genomic site primer. The peptides can also be used to immunise animals to
CC produce polyclonal or monoclonal antibodies with specificity for peptides
CC or proteins comprising the amino acid sequences. Antibodies are useful
CC for detecting the polypeptides in situ in cells or in vitro in cell
CC extracts. The present sequence represents a plant metabolism protein (or

XX OS Unidentified.
 XX PN WO200240694-A2.
 XX PD 23-MAY-2002.
 XX PD 16-NOV-2001; 2001WO-DK000763.
 XX PF 17-NOV-2000; 2000DK-00001730.
 XX PR 29-NOV-2000; 2000US-0253843P.
 XX PA (NOVO) NOVOZYMES AS.
 XX PI Vind J;
 XX PD WPI; 2002-471624/50.
 XX DR Heterologous production of taxane analogs related compound useful as
 XX PT anticancer agent involves cloning a full-length taxane synthesis pathway
 XX PT from a taxan-producing organism into a taxane-resistant host cell.
 XX PS Example 3; Fig 1; 60pp; English.
 XX CC The present invention describes the heterologous expression of taxane
 CC analogues, comprising cloning a DNA sequence comprising a taxane
 CC synthesis pathway, making a DNA construct in which the DNA sequence is
 CC under control of regulatory elements, introducing the DNA construct into
 CC a host cell, growing the host cell to produce taxane analogues, and
 CC recovering the required taxane from the culture medium. Also described:
 CC (1) an isolated DNA sequence comprising the taxane synthesis pathway,
 CC (preferably taxol synthesis pathway); (2) an expression vector comprising
 CC the isolated DNA sequence; and (3) a host cell comprising a taxane
 CC synthesis pathway derived from a taxane-producing organism, plant or
 CC tree, the taxane synthesis pathway is foreign to the host cell. The
 CC taxane analogues have cytostatic activity, and so can be used as
 CC anticancer agents. The present sequence represents a taxadiene synthase
 CC homologous protein, which is given in an example from the present
 CC invention
 XX SQ Sequence 868 AA;
 Query Match 25.8%; Score 1062.5; DB 5; Length 868;
 Best Local Similarity 32.3%; Pred. No. 2.1e-77;
 Matches 261; Conservative 148; Mismatches 339; Indels 61; Gaps 17;
 7 IASPLTKNRPAL---SAHTASTSHGQNTPTMLIIDTKERIQKQFNV---EISV 60
 79 LSSSTIVKREFFPGFWKDDILSTSSHKVAASDEKR-IETLISEIKNFRMFCMGYGETNP 137
 61 SSYDTAWAVPSPNPKPCPRCECLNLLNLDGSGWGLVNHHTHNPILLKDSLSST 120
 138 SAYDTAWARIFAVDGSDPHFFETVEWILQNLQKDGSGW-----EGFYFLAYDIRLAT 191
 121 LACIVALKRWNVGDEQINKGLSIESNLASATEKSQF-SPIGFDIIFPGLEAYAKNLIN 179
 192 LACIITLTLWRTGETQVKGIEFFRTQAGKWEADSHRPSGFEIVFPAMLKEAKILGLD 251
 180 LLSKQTDFFSLMLHKLRELOKRGCHSN---EMDGYLAVISEGLNLYDWNVKKYQMGNSV 236
 252 LPYDLFPFLQIIEKRAKLRIPTDVLIALPTLLYLSLEGLQEIWDQKIMLKQSDGSP 311
 237 FNSPSTAAAFINQFPCNLNLSLDDKFGNAVPTYPHDLFIRLSMDVTIERLGISSH 296
 312 LSSPASTAAVFMRTGNKKCLDLNVLNFKFGNVCHYPPLDLFERLWANDVTVERLGLDRH 371
 297 FVEIKNVLDIETVRCWVERD-----EQIFMDVVTCAIAFLRLRINGVEYSPDPIAETNE 351
 372 FKEEIKREALDYVYSHWDERGIGWARENFVDDIDTAMGLRILRLHGVNYSDDVLKTFRDE 431
 352 LALKDEYAA-----LETYHASTLYQIEDLSSGQILKSA-----DELKE 390
 432 ---NGEFFCFLGQTRGVTDMLNVNRCSHVSF-----PGETIMEEAKLCTERYLRNALEN 483

QY 391 IISTDSNRLSKLIHKEVENALKFPINTGLERINTERNIQLVNVNDTRILKITYHSSNISN 450
 DB 484 VDAFDKWAFFKNIRGEVEYVALKYPWHKSMPLREASYYIENYGPDDVWLKITYVMPIYSN 543
 QY 451 TDYLR LAVEDFYTCOSIYREELKGLERWVVENKLDQLPAROKTAYCYFSAATLSSPEL 510
 DB 544 EKYLEAKLDFNKVQSIHQTELQDLRRWKKSSGFTDLNFTREVRTEIYFSPASFIFEPEF 603
 QY 511 SDARISWAKNGILTVVDDFFDIGITIDELTNLIQCVEKNVVDVKDCCSEHVRILFLAL 570
 DB 604 SKREVTYTKTSNFTVILDDLYDAHGSLLDLKLFSTSVKRWDLSL-VDQMPQOMKICFVG 662
 QY 571 KDAICWIGDEAFKQWADVTSHTVITWLELMNSMLREALTWTRDAYVPTLNEYMENAYVSF 630
 DB 663 YNTFNDIAKEGRERQGRDVLGYIQNVKVLQLEAYYKEAEMSEAKYVPSFNEIENASVSI 722
 QY 631 ALGPVIVKPAIYFVGPKLSEIYESSEYHNLF-KIMSTQGRLLNDIHSFKREKFKGKLNNAV 689
 DB 723 ALGTVVLISALFTGEVLTDEVLSKIDRSRFTQLMGLTGRLVNDTKYQAEQGQGEV-AS 781
 QY 690 ALHLSNGESGKV-EEEVVEEMMMIKKEIKELMKLIFEENGSIIVPRACKDAFNMCHVLN 748
 DB 782 AIQCYMKDHPKUSEBEALQHVYSVMENALEELNR---EFVNNKIPDIYKRLVFETARIMQ 838
 QY 749 FFYANDDDGFT---GNTILDTVKDIIYNPL 774
 DB 839 LEYMQDGLTLSDHMEIKHEVKNCLFPQV 867
 RESULT 14
 ADM98968
 ID ADM98968 standard; protein; 868 AA.
 XX AC ADM98968;
 XX DT 01-JUL-2004 (first entry)
 XX DB Diterpene synthase polypeptide #6.
 XX KW Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;
 KW diterpene synthase; defence toxin; volatile defensive signal;
 KW pollinator attractant; photoprotectant; enzyme.
 XX OS Abies grandis.
 XX PN US2004072323-A1.
 XX PD 15-APR-2004.
 XX PF 07-JAN-2002; 2002US-00041018.
 XX PR 05-JAN-2001; 2001US-0259880P.
 XX PA (MATS/) MATSUDA S P T.
 XX PA (HART/) HART E A.
 XX PI Matsuda SPT, Hart EA;
 XX DR WPI; 2004-373921/35.
 XX PT New unicellular organisms comprising exogenous nucleic acids encoding a
 PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for
 PT producing diterpenes and diterpene precursors.
 XX PS Claim 31; SEQ ID NO 388; 38pp; English.
 XX CC The invention relates to a unicellular organism for producing a diterpene
 CC or diterpene precursor comprising an exogenous nucleic acid sequence
 CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
 CC promoter operable in the organism, and an exogenous nucleic acid sequence
 CC encoding a diterpene synthase under the control of a promoter operable in
 CC the organism. The invention also relates to methods of producing a

Mon Oct 18 07:21:11 2004

Qy		121	LACIVALKRNVGBDQINKGLSFTIESNIA GATEKSQP-SPIGFDIIIPGLLLEYAKNIIDIN	179
Db		192	LACIITLTLWRTGETVQVGKTESPRTQA GRKMEDEADSHRPSGFBIYVPFAMKEAKIIGLD	251
Qy		180	LLSKQTDFSLMLHKRELBOQRCHSN---EMDGYLAYISEGLNGLYDNMWNKKYQWKMGSV	236
Db		252	LPVDLPFLKQIEERAKLRIPDIVLYALPTTILYSLEGLOEIVEHEKIMKLQSDGSF	311
Qy		237	FNSPSATAAFINHQNPCCLNYLSNLDDKFONAVPTYVPHDLFIRLSMVDTERLGISHH	296
Db		312	LSSPASTAAVMFRMTGNKKCLDFLNFLVKFKFNHPVCHYPPLDFLFRLMAVDTVBERLIGDRH	371
Qy		297	FRVEIKNVLDETYRCWVERD-----EQIFMDVVTCALAFRI LLRRINGEVBSPPDLAEITNE	351
Db		372	FKEEIKGALDYVYSHWDERGICWARENPVDPIDDTAMGLRI RLIGHYVNVSSDVLTFRDE	431
Qy		352	LALMDEYAAA-----LETYHASHILIYQEDLSGGKOILKSA-----DFIKE	390
Db		432	--NGEFFCFLGTQGTORGVTDMLVNRCSHVSF----PGETIMBEAKLCTRYLRNALEN	483
Qy		391	IISTDSNRLSKLIHKEVENALKFPINTGLEIRINTERNIQLXVNDTRILKTTHSSNISLN	450
Db		484	VDAEPDKWAFKNIRGEVEYALKUPMHKSMPLREASYLENTGPDDVLWGKTYVMMPYLSN	543
Qy		451	TDYLRLAWEDEFYTQCSIYREELKGLERVWVNENKLDQLKPARQTAICYFSVAATLSSPEL	510
Db		544	EKYLELAKLDNFNKLOS IHQTELODLRRWKKSGGFTELNTFRERTVEITYFPSASFIFEPEF	603
Qy		511	SDARISWANGILITVVDFFDIGTTIDBLTNLTOCVEKNVVDVKDCSCSEHVRILTFLAL	570
Db		604	SKCREVYTKTSNFTVILDLYDAHGSLLDKFLTSTESKRVWLDSL-VDOQPKMKICTCFVGF	662
Qy		571	KDAICWI GDZFAFKWOARDVTSHVITWLELMNSMLREAITWRDAYVPTLINEYMENAYVSF	630
Db		663	YNTFNDIAKGRGRQGRDVLGVIONVKNVQLEAYTKEAESWSEAKYVPSFNFVIEINASVSI	722
Qy		631	ALGPITVKEPAIYVFGPKLSEBIEVESSEYNLP-KLMSTOGRLLINDIHSFKRFKEGKLNAY	689
Db		723	ALGTVLWLTSA LFTGVLGDVLSKIDRESRFTQLMGLTGRLVNDTPTYQAERGOGEV--AS	781
Qy		690	ALHLNSGSGKV-BEEVVEEMMMINKKRKELMKLKIIFRENGSIVPRACKDAFWNNCHVIN	748
Db		782	AIOCYMKDHPIKISEEALQHUVSYMVENALEELNR---EFVNKKIPDIYKELVFETARI MQ	838
Qy		749	PFYANDDGFT---GNTILDTVKDIIYNPL	774
Db		839	LFYMOGDGLITLSDHMEIKRHVCNKCIFOPV	867

Search completed: October 17, 2004, 08:45:53
Job time : 212.188 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 17:27:39 ; Search time 8680.1 Seconds
(without alignments)
11721.026 Million cell updates/sec

Title: US-10-041-018-361

Perfect score: 2792

Sequence: 1 cccactcatctttatcaac.....aaaaaaaaaaaaaaaaaaaaa 2792

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	457.2	16.4	725	5	BQ992296 QGF25D22.
2	393.6	14.1	510	1	AU292191 AU292191.
3	337	12.1	665	1	AI485628 EST243949
4	316.8	11.3	632	2	BE434899 EST405977
5	310.8	11.1	843	7	CO079063 GR_Ea41K
6	305.2	10.9	590	1	AI485623 EST243944
7	239.4	8.6	753	2	BE658119 GM700004B
8	233	8.3	821	6	CB894537 EST647329
9	219	7.8	500	5	BQ104995 S90295.e
10	207.2	7.4	367	6	CA824186 R37G11 tw
11	205.6	7.4	1077	5	BX838565 BX838565
12	195.8	7.0	543	6	CF037118 QCG40h09.
13	190.4	6.8	619	6	CF197159 Ir64v11.5
14	189.6	6.8	567	6	CF036221 QCG29C10.
15	185	6.6	515	6	CF037618 QCG9506.y
16	183.6	6.6	596	6	CA452452 Rxo-2_D07
17	174.6	6.3	582	6	CF036148 QCG28B11.
18	174.6	6.3	582	6	CF036916 QCG38G06.
19	172.8	6.2	439	1	AI940878 sb79b02.y
20	172.2	6.2	558	1	AU252376 AU252376
21	171.6	6.1	443	4	B1949839 HVSMEL001
22	164.6	5.9	808	6	CB621697 OSIIea07M
23	164.4	5.9	646	7	CF418783 USDA-FP 1
24	164.4	5.9	685	2	AW738064 EST339491

25	163.2	5.8	529	6	CA914596	PCSC09424
26	160.6	5.8	197	1	AU294693	AU294693
27	158	5.7	526	6	CB899647	PCBP02741
28	157.4	5.6	788	6	CB623234	OSIIra10N
29	155.4	5.6	575	4	BM318434	PI1_13_E1
30	154.8	5.5	547	2	AW617347	EST323758
31	154	5.5	699	5	BQ766175	BBR008_SQ
32	152.8	5.5	825	5	CB623957	OSIIra12B
33	146.2	5.2	555	2	AW617193	EST323604
34	145.6	5.2	618	2	AW616201	EST307240
35	145.6	5.2	616	2	AW616591	EST323002
36	145.6	5.2	620	2	AW616079	EST296838
37	145.6	5.2	625	2	BE353211	EST400343
38	145.6	5.2	642	2	AW616634	EST323045
39	145.6	5.2	705	4	BI928166	EST348055
40	143.8	5.2	614	2	AW616902	EST323313
41	142.6	5.1	689	2	BE585476	EST46P9P6
42	142.4	5.1	556	4	BG126452	EST472098
43	141.4	5.1	452	4	BM501430	PAC000000
44	140.2	5.0	718	4	BQ837019	Zm08_1060
45	139.2	5.0	531	6	CA646242	wreln.pk0

ALIGNMENTS

RESULT 1
BQ992296
LOCUS BQ992296 725 bp mRNA linear EST 21-AUG-2002
DEFINITION QGF25D22.yg.ab1 QG EFGHJ lettuce serriola Lactuca sativa cDNA clone
QGF25D22, mRNA sequence.
ACCESSION BQ992296
VERSION BQ992296.1 GI:22411831
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichorieae; Lactuca.
1 (bases 1 to 725)
Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,
Ellison,P., Kolkman,J., Slabaugh,M.S., Livingson,K., Zhou,Y.,
Lai,Z., Church,S., Jackson,L. and Bradford,K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compgenomics.ucdavis.edu/
Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmunsen Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atg.c.org [michelmore@vegmail.ucdavis.edu]
singleton, see http://cgdb.ucdavis.edu/ for details.
Plate: QGF25 row: D column: 22.

FEATURES
source
1. 725
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="L.serriola"
/db_xref="taxon:4236"
/clone="QGF25D22"
/lab_host="E.coli"
/clone_lib="QG EFGHJ lettuce serriola"
/note="Vector: pBRCNAsFAB: The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize

[illegible][illegible]

```

Db      1  TTAAGTTTCCTTCATCGCAATATAAAACGGGTAGCCCAACAGGAGAAAATATAAAGCACT 60
QY      1465 ACAATGTAGACATACAGAAATCTGAAACTACATATACATCATCAATATATAGTAACA 1524
Db      61  ATGATGTAGATAATAAAGAGTCTGAAAGACTTCATATATAGTTTCATCAATATTTGGCAACA 120
QY      1525 CTGATTTACCTAAGTTGGCTGTGTGAAGATTCTACACCTGCCAATCTATTTATCGTGAAG 1584
Db      121 AAGATTTCTTAACCTTGGCGGTAGAAAGACTTCAACCTTTGCCAATCTATCCACCGCAAG 180
QY      1585 AATTAAGGTCTGTGAAAGGTGGGTGAGAGATAAGTTGGACCAAGCTCAAGTTTGTCTA 1644
Db      181 AATTAAGCACTTGAAGAGTGTGTAACGCAAAACAGATTGGACAAAGCTGAAGTTCTGTCA 240
QY      1645 GGCATAAGACCGCTACTGTTATTTCTGTTGCTGCAACACTTTCTGCTCCGGAATTA 1704
Db      241 GAGAGAGATCTGCATACCTGCTACTTTCTGCTGCAGCCCAATTTTCAACCTGGAATAT 300
QY      1705 CAGATCGCGTATTTTCATGGGCCAAAATGGCAATTAATCACTACAGTAGTTGATGACTTT 1764
Db      301 CTGATGCCGCACTGATGGCCCAAGATGGTGTACTTACTACAGTGAATGATGATTTCT 360
QY      1765 TTGATACGGTGTGTAACAATGCAATGCAACCACTGATTCATGATGTTGAAAGATGGA 1824
Db      361 TTGATGTAGAGGTTCTATGGAAGAATTGAACCACTAATTCGTGTTTAAAGAGTGG 420
QY      1825 ATGTAGATGTGCAAGGATTTGTTTCAGAGCATGTTCCGATTTTATTTTATAGCAATTA 1884
Db      421 ATGTAGATGTGACATGATGCTGTTTCGAGAGAGTAGGCATTTATTTTCGGCGCTTC 480
QY      1885 AAGATCAATCTCTGATGGAGATGAAGCTTTTAAATGGCAAGCGCGATGTAACATA 1944
Db      481 ACAGTACTAATAGTGAGATTGAGACAAAGCATCTAAGTGGCAAGCAGCTAGTGTACAA 540
QY      1945 GCATGTTATTCAACTGGTGGAACTAATGATAGTATGTTGAGAGAGCTATATGGA 2004
Db      541 GACACATACTGATATTTGGTTAAATCTACTGAATGCTATGTTAAGAGAAGCTGAATGG 600
QY      2005 CAAGAGATGTTATGTCACAACTAATTAAGAA 2036
Db      601 CTAAGATATGTCGGTGCCATCATTTGGACAAA 632

```

```

RESULT 5
LOCUS   CO079063
DEFINITION GR_Ea41K03.r GR_Ea Gossypium raimondii cDNA clone GR_Ea41K03 3', mRNA sequence.
ACCESSION CO079063
VERSION   CO079063.1 GI:48748544
KEYWORDS EST.
SOURCE    Gossypium raimondii
ORGANISM  Gossypium raimondii
REFERENCE 1 (bases 1 to 843)
AUTHORS   Kim,H., Yu,Y., Kudrna,D., Hatfield,J., Stum,D., Mueller,C., Udall,J.A., Rapp,R.A., Wendel,J.F., Rao,K., Soderlund,C. and Wing,R.A.
TITLE      Global assembly of Cotton ESTs
JOURNAL    Unpublished (2004)
COMMENT    Contact: Rod A. Wing
            Arizona Genomics Institute
            The University of Arizona
            Forbes Building Room 303, Tucson, AZ, 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 41 row: K column: 03.
            Location/Qualifiers
                1..843
                /organism="Gossypium raimondii"

```

```

/mol_type="mRNA"
/db_xref="taxon:29730"
/clone="GR_Ea41K03"
/tissue_type="whole seedlings"
/dev_stage="first true leaves"
/lab_host="DH10B"
/clone_lib="GR_Ea"
/note="Vector: pCMV.SPORT-6.1; Site 1: NotI; Site 2: EcoRV; Library made by Invitrogen with RNA supplied by Wendle lab. Directional clones into NotI-EV. Colonies plated/picked by AGI. More glycerol clones held in -80."

ORIGIN
Query Match      11.1%; Score 310.8; DB 7; Length 843;
Best Local Similarity 63.8%; Pred. No. 6e-61;
Matches 537; Conservative 0; Mismatches 272; Indels 33; Gaps 3;

QY      939 CTTAATTTAAATTCACCTTTTGGACAAAGTTTGGTAATGCGATGCCCAACAGTTTATCCT 998
Db      1  CTTCAATATCTCACTGCACCTTTAGACCGGTTTGAGATGCGAGTTCCACACCTTCATCCT 60
QY      999 CATGATTTATTTATCCGACTTTCTATGGTTGACACAAATTTGAAAGATTAGGAATTTCCAC 1058
Db      61  TTCCATGTTTTCGCCGACTTTGTATGTTGGAGACTGTTGAAAGTTTGGAAATTTGTCAG 120
QY      1059 CATTTCAGAGTGGAAATTTAAATGTTTATAGATGAAACATACAGATGTTGGGTGGAACCA 1118
Db      121 CATTTCAGGAGAAATTCACAGTGTATTAGATGAAACATATCGATTTGGCTACAGGG 180
QY      1119 GATGACAAATATTCATGATGTTTAAACATGTGCTTTAGCCCTTCGGTATTATTAAGATC 1178
Db      181 GAGGAGGAGATATCTTGATCTGCGACATGTGCAATGCTTTCAGAACACATACCTGTA 240
QY      1179 AATGGTATGAAGTTTCCCGACATCCATTGGCTGAAATTTACTAA 1222
Db      241 AAGCGTTATGATGTTTCCCTCAGAAGCCTTAGCTGGATTTGCTGGAAGAACACTTCTCAAT 300
QY      1223 -----TGAATTAGCTTTGAAAGCAATATGCAAGCTCTTTGAAACATATCATCGTGCACA- 1276
Db      301 TCACITGGAGGATATTTGAGGATTTGGATGCTGCTGTTGAGCTATTTAGGGCTTCACAG 360
QY      1277 --TATATTATACCAAGAGGATTTATCTCTGGAAGAAACAAATCTTGAAGTCAAGTGAATTC 1334
Db      361 ATGATCATACATCCAAATGAACAACTCTCGGAGAAACACATCTCATGGAAGCAAGTCATTC 420
QY      1335 CTCGAAGATATATATCCACTGATTCGAACAGGCTTTCTAAATTAATTTACACA----- 1386
Db      421 TTGAAACAGGAATTTATCCAAACACTTCTAAATGTGCATATAAACAATAAATAATATATG 480
QY      1387 -AAGAGGTGGAATAATGCTCTTAAGTTCCCTATCAATACCGGTTTAGAAGCGCATAAACACT 1445
Db      481 CAGAAGTGAAATGATGCCCTTGAGTTTCCCATTAATGCAAGTTTGGAGCGTTAGTATAC 540
QY      1446 AGACGAAATATACAGCTTTTACAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1505
Db      541 AGGAGAAACATAGTGAACATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY      1506 TCATCAAAATATAGTAAACACTGATTAAGGTTGCGCTGTTGGAAGATTTCTACACTGC 1565
Db      601 TCTTTGAGTATTGGCAACAAAGATTTCTTAAGCTAGCAGTGGAGGACTTCAATGCGCTGT 660
QY      1566 CAATCTATTATCTCGTGAAGAAATTTAAAGGTCTTTGAAAGGTGGGTGGGTAGAGATAAGTTG 1625
Db      661 CAATCTATATATCTGTGAAGAACTCAAAACACTCGAGAGGTGGGTTCAGAAAAGCGATTG 720
QY      1626 GACCACTCAAGTTTCTAGGCAAGACCGCTACTGTTATTTCTGTTGCTGCAACA 1685
Db      721 GACAGCTGAAGTTTTCAGGCGAAGAACTGGCATACTGTTACTTTTCTGAGCTGCAACA 780
QY      1686 CTTTGTCTCCGAAATTTATCAGATGCGGATTTTTCATGGCCAAAATGCAATTAATTA 1745
Db      781 CTATGCCCTCTGAATTAATCTGATGCTCGCTTAACATGGCTAAATGCGCTGCTTACT 840

```

```

QY      1746 AC 1747
Db      841 AC 842

RESULT 6
AI485623
LOCUS   EST243944 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
DEFINITION
ACCESSION AI485623
VERSION   EST18-MAY-2001
KEYWORDS  CLD6L20, mRNA sequence.
SOURCE    Lycopersicon esculentum (tomato)
ORGANISM  Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 590)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.W., Venter,J.C.,
Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
Location/Qualifiers
1..590
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/cloned="CLD6L20"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRF"
/clone_lib="tomato ovary, TAMU"
/note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:
XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector lambda Zap II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."

FEATURES
source
Query Match 10.9%; Score 305.2; DB 1; Length 590;
Best Local Similarity 69.8%; Pred. No. 1.1e-59;
Matches 412; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY      1489 TGAACAACATATACATCAATATATAGTAACACTGATTACTAGTCTAGGTTGGCTGTG 1548
Db      1 TGAACAACATATATAGTTCATCAAAATTTTGGCAACAAGATTTCTTAACCCCTGGCGGTAG 60

QY      1549 AAGATTCTTACACCTGCGCAATCTATTATTCGTGAGAAATTAAGGTTCTTGAAGTGGG 1608
Db      61 AAGACTTCAACTTTGGCAATCTATCCACCGCAAGAAATTAACAACATGAGAGGTGGT 120

QY      1609 TGGTAGAATAAGTTGGACGAGCTCAAGTTTGGTCAGCAAAAGACCCGCTACTGTTATT 1668
Db      121 TAACGCAAAACAGATTGGACAAGCTGAAGTTTCGTGAGAGAGATCTGCATACTGCTACT 180

QY      1669 TCTCTGTTGCTGCAACACTTTGCTGCCGAATATACAGATGCGCGTATTTTCATGGGCCA 1728
Db      181 TTTCTGCTGCGCAACAAATTTTCAACCTGAACTATCTGATGCGCGCATGTTCATGGGCCA 240

QY      1729 AAAATGGCATATTACTACAGTATGATGACTTTTTCATATCGGTGGTCAATCGATG 1788
Db      241 AGAATGGTGTACTTACTACAGTATGATGATTTCTTTGATGTAGGAGGTTCTATGGAAG 300

QY      1789 AATTGACCAACTGATTCAATGTGTGTGAAAAATGGAATGTAGATGTCGCAAGGATTGTT 1848

```

```

Db      301 AATTGAACAACCTAATTTCTGTTTAAAGAAGTGGAGTAGATGTACCACTGATGCT 360
QY      1849 GTTCAGAGCATGTTCCGATTTTATTTTATTTAGCAATTAAGAATCAATCTCTTGGATTGGAG 1908
Db      361 GTTCGAGAGAGTAGGCAATTTATTTTCGGCGCTTCACAGTACTATAAGTGAGATTGGAG 420
QY      1909 ATGAAGCTTTAAATGGCAAGCGCCGATGTAACTAGCCATGTTTAACTTGGTTGG 1968
Db      421 ACAAAGCATCTAAGTGGCAAGCACGCTAGTGTGTACAGACACACATACTGATATTTGGTTAA 480
QY      1969 AACTAATCAATAGTATGTTTGAAGAAGCTATATGACAAAGAGATGCTTATGTGCCAAT 2028
Db      481 ATCTACTGAATGCTATGTTTAAAGAAGCTGATGGCTTAAGATATGTCTGGTGCATCAT 540
QY      2029 TAAATGAATATATGAAAACGCTTACGTCTCATTTGCAATTAGCCCGATT 2078
Db      541 TGGACAAATATATGCGCAACGCTTATGTATCATTTTGCCTAGGACCAATT 590

RESULT 7
BE658119/c
LOCUS   BE658119
DEFINITION BE658119 753 bp mRNA linear EST 24-MAY-2001
GM700004B20H10 Gm-r1070 Glycine max cDNA clone Gm-r1070-1532 3',
mRNA sequence.
ACCESSION BE658119
VERSION   BE658119.1 GI:9984011
KEYWORDS  Glycine max (soybean)
SOURCE    Glycine max
ORGANISM  Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 753)
Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Corvelli,V.,
Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished (1999)
Other ESTs: AI940878 corresponding to Gm-ci010-1156 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for
Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional
Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134. For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or info@genome
systems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
Location/Qualifiers
1..753
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="Gm-r1070-1532"
/clone_lib="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, reracked
set of 9,216 clones selected from cDNA libraries from
various tissues and stages of development of soybean that
represent 2,639 sequences from immature cotyledons, 1,770
from immature seed coats, 3,938 from flowers, and 869
from young pods. The 5' ESTs of the source clones from
the different libraries was used to select singletons, or
a representative of each contig, which were reracked to
form library Gm-r1070. The cDNA clones of the reracked
Gm-r1070 library were then sequenced at the 3' end. The
contig analysis to select unique genes was performed by

```

the laboratory of Ernest Retzl, Center for Computational Genomics and Bioinformatics, University of Minnesota, <http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html>. Reracking was performed by Genome Systems, St. Louis, <http://www.genomesystems.com>, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, <http://www.life.uiuc.edu/biotech/keck.html>. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'.

ORIGIN

Query Match 8.6%; Score 239.4; DB 2; Length 753;
Best Local Similarity 60.5%; Pred. No. 2e-44;
Matches 369; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 1892 AATCTCTGCGATGGAGATGAAGCTTTTAAATGCGAAGCGCGGATGTAAGTACAGCCATGT 1951
DB 748 AGTTNNTGANNNNAGANNNTGCTCANNAGCAGGCGCAATGTGAATAACATGT 689

QY 1952 TATTCAAACTTGGTGGAACTAATGAATGATATGTTTCAGAGAAAGCTATATGGAACAAGAGA 2011
DB 688 TATCAAAATTTGNNNGNAITTTGCTCCAGTCAATGTTTAGAGAGCTGAGTGGTTGAGAAC 629

QY 2012 TGGTTATGCGCAACAATTAATGAATATATGGAAGCGTGTAGTGTCAATTTGCATTAGG 2071
DB 628 CAAGACTGTGCCAACAAATTTGCTACTATATGGAAGATGCAATATCATTTGNCNTAGG 569

QY 2072 CCGGATGTCAAGCGCGTATTTACTTTGTTGGGGCCCAATATTCAGAGAGAGATTTGTGA 2131
DB 568 ACCAATGTCTTCCGGCCNCTATNNGTTGGANNNAAGCTTTTCAGATGAATGCACAGA 509

QY 2132 AAGCTCTGAATATCATAAATCTATTTAAGCTTAATGAGCAGCAGCGGTGCACTTCTAAACGA 2191
DB 508 AATCATGAGTTGAACTATCTCTATAAGCTCATGACACATGTGGCGCTTCTCTTAATGA 449

QY 2192 TATCCATAGTTCAGAGGGAATTTAAGAGGCAATTAAGCGCGGTAGCATTTGCATTT 2251
DB 448 CATTTCACGTTTAAAGAGAGAAATCCGAGGAAGGAAATGGAATGTGTGCTGCGTAT 389

QY 2252 GAGTAACGAGAAAGTGGGAAGTGGGAAGAGAGTGTGTGGAGAGATGATGATGAT 2311
DB 388 TGCTCTATGCGAATGGAGTTATTTACTGCGAAGATGCCACGAGAAATGAGGGTATGC 329

QY 2312 TAAAAACAAGAGAAAGAAATTAATGAATTAATTTTGAAGAAATGGTAGCATTTGTTCC 2371
DB 328 TGAAGAAAGAGAGAGAACTCTCTGAGATTAAATTTGAGAAAGAAAGAGTGTAGTTCC 269

QY 2372 TAGAGCTTGAAGATGCATTTTGAACATGTGTACGTTGTAATTTTTCACGCAA 2431
DB 268 CAGAGAGTGCAGAGATTTGTTTGGAAATGATAAAAGTGTGCACCTGTTTACATGAA 209

QY 2432 CGATGCGGTTTACTGGAACACGATTTCTTGATCTGTGAAGACATCATTTTACAAACC 2491
DB 208 GGAATGATTTACTTTCACACAGATGCACTCTAGTGAATGCACTATTAAAAACC 149

QY 2492 GTTGGTGCCTT 2501
DB 148 TGTCACTCCTT 139

RESULT 8

CB894537
LOCUS
DEFINITION
EST647329 HOGA Medicago truncatula cDNA clone HOGA-31D19, mRNA
sequence.
ACCESSION
CB894537
VERSION
CB894537.1 GI:30101706
KEYWORDS
EST.

SOURCE
ORGANISM

Medicago truncatula (barrel medic)

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

1 (bases 1 to 821)

REFERENCE

AUTHORS

Hahn, M.G., Ojane-Reuhs, T., Samac, D., Town, C.D., Van Aken, S.,

Uterback, T., Cho, J. and Fraser, C.M.

ESTs from roots of Medicago truncatula treated with

oligogalacturonides of DP 6-20

Unpublished (2001)

Contact: Michael G. Hahn

Complex Carbohydrate Research Center

University of Georgia

220 Riverbend Road, Athens, GA 30602-4712, USA

Tel: 706-542-4457

Fax: 706-542-4412

Email: hahn@ccrc.uga.edu

TIGR sequence name: MTMDU22TK

More information is available at: www.medicago.org

Seq primer: SKmod (CTA GAA CTA gtg gat CC).

FEATURES

Location/Qualifiers

1..821

/organism="Medicago truncatula"

/mol_type="mRNA"

/cultivar="A17"

/db_xref="taxon:3880"

/clone="HOGA-31D19"

/tissue_type="3 day old seedling roots"

/dev_stage="24 hours after treatment in the dark at 26 C

with 0.5 mg/ml oligogalacturonides (DP 6-20) in the

presence of 100 ug/ml Gentamicin"

/lab_host="XLOLR"

/clone_lib="HOGA"

/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in SOLR cells."

ORIGIN

Query Match 8.3%; Score 233; DB 6; Length 821;
Best Local Similarity 62.9%; Pred. No. 6.3e-43;
Matches 403; Conservative 0; Mismatches 220; Indels 18; Gaps 2;

QY 1381 TTCACAAAGAGGTGGAAATGCTTTAAGTTCCTATCAATACCGGTTTGAACGCAATAA 1440
DB 180 TTGATAATGAGTCAATGATGTTCTTAAGTTTCCACATCATGCGAATTTAGAGCGTTGT 239

QY 1441 ACACATGAGCAATATACAGCTTTTACAATGTACACATACAGAAATCTGAAACATCAT 1500
DB 240 TAAACGAGAGATCAGTGGAGCATTAATGCGATGAACAGAGATTTGAAACATCCT 299

QY 1501 ATCACATCATCAATATATTAGTAAACATGATTACCTAAGGTTGGCTGTTGAAGATTTCTACA 1560
DB 300 ACAGATCATGCATCTTGCACCAAGAAATCTGAGCTAGTAGCAGTTGAAGACTTCAACC 359

QY 1561 CTGCGCAATCTATTATTCGTGAAGAATTAAGAGTCTTGAAGGTGGTGGTGTAGAGATA 1620
DB 360 TCTGCGCAATCAATACAAAATAAAGAGCTGAAACAACTTTCAGGTGGATTTGTGAAGCA 419

QY 1621 AGTTGACAGCTCAAGTTTGTAGCAAAAGACCGCTACTCTGTTCTCTGTTGCTG 1680
DB 420 AACTGACAACTAGACTTTTGGAGGCGAGAACTGGCATACTGTTACTTCTCAGTGCAG 479

QY 1681 CAACACTTTGCTCTCCGGAATTTATCAGATGCGGTATTTTCATCGGCGCAAAATGGCATAT 1740
DB 480 CTACTCTTTATTTCCCGCACTTCTCTGATGCTGCGATATCTTGGCGCAAAATGGGTAC 539

QY 1741 TAACATAGTAGTTGATGACTTTTTTGTATATCGTGTGTACATCGATGAATTCAGCAACC 1800

```

Db      540 TCACAACAGCTGCTGATGATTTCTTGATGTCGGGAGTTCTGAAGAAGAGCAAGTGAACC 599
QY      1801 TGAATCAATGCTGTGA AAA-----TGAATGATAGATGTGACAGGA 1843
Db      600 TTATTCACAGTGAAGAGATGATATTCCTAGTATTTAGTGGGATGTAGATGTCAATACTGT 659
QY      1844 TTGTTGTTCCAGAGCATGTTCCGATTTTATTTTATTTAGCATTTAAAAGATGCAATCTGTGGAT 1903
Db      660 TTTCTGTTCTGAGGCGGTAAAGATATATTTCTGAGTTCGTAGCACAAATTTGGAGAT 719
QY      1904 TGGAGATGAAGCTTTAAATGGCAAGCGCGCGATGTAATAGCCATGTTAT-TCAAAATT 1962
Db      720 TGGAGAGAAATCTGTTAGCGGCAAGCGCGCAATGTGAAAGACAATGTTATCAAAAATTT 779
QY      1963 GGTTCGAACTAATGATAGTATGTTGAGAGAGCTATATGG 2003
Db      780 GGTTAGATTGATGCGCTCTATGTTTACAGAGCTGAGTGG 820

RESULT 9
LOCUS   BQ104995          500 bp      mRNA      linear      EST 16-APR-2002
DEFINITION   S90295.e Rose Petals (Golden Gate) Lambda Zap Express Library Rosa
              hybrid cultivar cDNA clone gg0295.e 5', mRNA sequence.
ACCESSION   BQ104995
VERSION     BQ104995.1  GI:20154657
KEYWORDS    EST.
SOURCE      Rosa hybrid cultivar
            Rosa hybrid cultivar
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Rosales; Rosaceae; Rosoideae; Rosa.
REFERENCE   1 (bases 1 to 500)
            Gureman,I., Shalit,M., Menda,N., Piestun,D., Dafny-Yelin,M.,
            Shalev,G., Bar,E., Davydov,O., Ovadis,M., Emanuel,M., Wang,J.,
            Adam,Z., Fichersky,E., Lewinsohn,E., Zamir,D., Vainstein,A. and
            Weiss,D.
TITLE       Rose Scent: Genomics Approach to Discovering Novel Floral
            Fragrance-Related Genes
JOURNAL     Plant Cell 14 (10), 2325-2338 (2002)
COMMENT     Contact: Naama Menda
            Petal Genomics
            Faculty of Agricultural, Food and Environmental Quality Sciences,
            The Hebrew University of Jerusalem
            P.O. Box 12, Rehovot, 76100, Israel
            Tel: 972 8 9489 683
            Fax: 972 8 9468 265
            Email: shanama@agri.huji.ac.il
            Seq primer: T3 forward.
            Location/Qualifiers
                1..500
                /organism="Rosa hybrid cultivar"
                /mol_type="mRNA"
                /strain="Golden Gate"
                /db_xref="taxon:128735"
                /clone="gg0295.e"
                /tissue_type="Petals"
                /dev_stage="Young open flower at stage four"
                /clone_lib="Rose Petals (Golden Gate) Lambda Zap Express
                Library"
                /note="Vector: pBKMV; Site_1: EcoRI; Site_2: XhoI"

FEATURES             source
    Query Match          7.8%; Score 219; DB 5; Length 500;
    Best Local Similarity 65.8%; Pred.No.1e-39;
    Matches 318; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY      1922 ATGGCAAGCGGCGATGTAAGTCCATGTTATTCAAACTGTTGGAACTAATGAATAG 1981
Db      10  AAGCAAGCGGCGAGTGTGACAAATCAGTCATTGAGATTGTTGGATTGCTCAAGGC 69
QY      1982 TATGTTGAGAGAGCTATATGGACAAGAGATGCTTTATGTGCCAACATTAAATGAATAT 2041

```

```

Db      70  TATGCTGAGGAGAGCTAGTGTGTTGATAAACAGTCAGCGCCACCAATGGAGGATACAT 129
QY      2042 GGAACACGCTTACGTGTCATTTGCAATTAGCCCGGATTTGCAAGCCGGCTATTACTTTGT 2101
Db      130 CGAAATGCGTACGTCTCATTTTGCCTTGGACCTAATGTCTTCAAATCTCTATTGTTGT 189
QY      2102 GGGGCCCAATTTATCAGAGGAGATTTGTTGAAGCTCTGAATATCATATCTATTAAAGT 2161
Db      190 AGGGCTTAAGCTCTCAGAGGATGCTGTACGGAGTTTCAAGATACTCCATCTCTATAGACT 249
QY      2162 AATGAGCAGCGCAGGCTCGACCTTTAAACGATATCCATAGCTTCAAGAGGGAATTTAAGGA 2221
Db      250 AATGAGCACTTTTGGCGCTCTTCTCAATGATATGCAAGCTATAGAGGGAAGCTGCCGA 309
QY      2222 AGGCAAAATTAACGCGGTAGCATTTGCATTTAGTAAACGAGAAAGTGGGAAAGTGGAGA 2281
Db      310 AGGGAAGCTAAATGCTGTAAACATTTGCCCATGTTTATGGCAATGTGTAGTTAGCGAGA 369
QY      2282 AGAGGTTGTGGAGGAGATGATGATGATGATTAATAACACAGAGAGAAAGATTAATGAAT 2341
Db      370 AGAGGCCATCAATGAGATGAAGAGCATTTATCACCACTAAGAGGAGAGAACTGCAAGACT 429
QY      2342 AATTTTCAAGAAAATGTTAGCATTTGTTCTTAGAGCTTGTAAAGATGCAATTTTGGAACT 2401
Db      430 AGTTTTCAGGACACAGGATAGTGTAGTTCCAGAGCTTGCAGAGATCTGTTTGGAACT 489
QY      2402 GTG 2404
Db      490 GGG 492

RESULT 10
LOCUS   CA824186          367 bp      mRNA      linear      EST 28-FEB-2003
DEFINITION   R37G11 two-month-old roots from clone 'Beaupre', Populus balsamifera
              subsp. trichocarpa x Populus deltoides cDNA 5', mRNA sequence.
ACCESSION   CA824186
VERSION     CA824186.1  GI:28607735
KEYWORDS    EST.
SOURCE      Populus balsamifera subsp. trichocarpa x Populus deltoides
            Populus balsamifera subsp. trichocarpa x Populus deltoides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Malpighiales; Salicaceae; Populus.
            1 (bases 1 to 367)
            Kohler,A., Delaruelle,C., Martin,D., Encelet,N. and Martin,F.
            The poplar root transcriptome: analysis of 7000 expressed sequence
            tags
JOURNAL     FEBS Lett. 542 (1-3), 37-41 (2003)
COMMENT     Contact: Martin FM
            Equipe de Microbiologie Forestiere
            Institut National de la Recherche Agronomique
            Centre INRA de Nancy, 54280 Champenoux, France
            Tel: +33 383 39 40 80
            Fax: +33 383 39 40 69
            Email: fmartin@nancy.inra.fr
            Insert Length: 367 Std Error: 0.00
            Seq primer: Fomat 5' AAGCGCGCCATTGTGTGTGTTACCC.
            Location/Qualifiers
                1..367
                /organism="Populus balsamifera subsp. trichocarpa x
                Populus deltoides"
                /mol_type="mRNA"
                /cultivar="Beaupre"
                /db_xref="taxon:3695"
                /dev_stage="two-month-old"
                /clone_lib="two-month-old roots from clone 'Beaupre'"
                /note="Organ: root; Vector: pRiplex2; cDNA library of
                two-month-old roots from Populus trichocarpa Torr.& Gray
                x deltoides Bartr. Ex Marshall (clone 'Beaupre'). The cDNA
                library was constructed from 1 ug of total RNA using the
                SMART cDNA synthesis kit (Clontech, Palo Alto, CA, USA)

FEATURES             source
    source

```

according to the manufacturer's instructions. The resulting cDNA was packed into lambda phages using the Gigapack III Gold packaging kit (Stratagene, La Jolla, CA). The pTriplex2 phagemid clones in Escherichia coli were obtained by using the mass in vivo excision protocol according to the manufacturer's instructions (Clontech)."

ORIGIN	Query Match	7.4%; Score 207.2; DB 6; Length 367;
	Best Local Similarity	73.9%; Pred. No. 5.2e-37;
	Matches	263; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
QY	1473	GACATACAGAAATCTGAAACTACATATCACTCAATCAATATAGTAACTGATTAC 1532
DB	10	GATGATACAGAGATTTGAAACTTCGTATCGTTGTTCCACTATTTGGTAAACCGAATTTT 69
QY	1533	CTAAGGTTCGGCTGTTGAAGATTTTACACCTGCAATCTATTTATCGTGAAGAAATTAATA 1592
DB	70	CTCAATTCGCTGTAGAGACTTCAATATCTGCCAATCAATACAGCGTGAAGAAATTAATA 129
QY	1593	GGTCTTGAAGGTGGTGTAGAGATAAGTTGGACCACTCAAGTTTGGTGGCAAAAG 1652
DB	130	CATATCGAGGTGGGTTGTAGAGAGAGACTGGACAAGCTAAAGTTTGTCTAGGCGAAG 189
QY	1653	ACGCGCTACTGTTATTTCTGTTGCTGCAACACTTTTGGTCTCCGAAATTTATCAGATCG 1712
DB	190	GAGGCTACTGTTACTTCTGCTGCGAGCAACTCTCTTGGTCCAGAACTATCTGAATGA 249
QY	1713	CGTATTTTCATGGCCAAAATGSCATATTAATACAGTATGATGACTTTTTTGTATATC 1772
DB	250	CGCATGTCATGCGCAAAAATGTTGTTCTTACTACTGTTGTTGATGACTTCTTTGATGTT 309
QY	1773	GGTGTACAAATCATGAATGACCAACCTGATTCATGTTTCAAAAATGGAATGT 1828
DB	310	GGTGGTCTCTGAAGAAGAACTGGTAAACCTTTATGAATTTGATGAGAGGTGGGAATG 365

RESULT 11
 BX838565
 LOCUS
 DEFINITION
 Arabidopsis thaliana Flowers and buds Col-0 Arabidopsis
 thaliana cDNA clone GSLTFB67ZB05 5PRIM, mRNA sequence.
 BX838565
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 1 (bases 1 to 1077)
 Castelli, V., Aury, J.M., Jaillon, O., Wincker, P., Clepet, C.,
 Menard, M., Cruaud, C., Quetier, F., Scarpelli, C., Schachter, V.,
 Temple, G., Caboche, M., Weissenbach, J. and Salanoubat, M.
 Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences:
 A Combined Approach to Evaluate and Improve Arabidopsis Genome
 Annotation
 Unpublished (2004)
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 The sequences are based on single pass reads.
 Life technologies (a division of Invitrogen) members carried out
 full-length libraries construction : Temple G.
 Genoscope members carried out sequencing and annotation : Castelli
 V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C.,
 Schachter V., Weissenbach J., Salanoubat M.
 URGV INRA : Clepet C., Caboche M.
 Annotation is based on the June 2003 version of the Arabidopsis
 genome released by MIPS (Munich Information center for Protein
 Sequences).
 http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/EST

http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.

FEATURES	Location/Qualifiers	source
	1..1077	
	/organism="Arabidopsis thaliana"	
	/mol_type="mRNA"	
	/ecotype="Col-0"	
	/db_xref="taxon:3702"	
	/clone="GSLTFB67ZB05"	
	/tissue_type="Flowers and buds"	
	/clone_lib="Arabidopsis thaliana Flowers and buds Col-0"	
ORIGIN	Query Match	7.4%; Score 205.6; DB 5; Length 1077;
	Best Local Similarity	54.7%; Pred. No. 1.5e-36;
	Matches	527; Conservative 0; Mismatches 414; Indels 23; Gaps 5;
QY	280	ATCCCACTAATCTGATCATTTGATACAAACCAAGACGGATCCAAAACAGTTTAAAAATG 339
DB	41	AGCTAACATGTAGAGCTTTTGAGCAACAAAGAGAGAGATTAGGAAGATGTTGGAGAAAG 100
QY	340	TAGAAATTTCTGTTCTTCATATGACACAGCATGGGTAGCCATGGTCCCTTCTCCAAACT 399
DB	101	TGAGCTTTCTGTTTCGGCTAGATACATCTAGTTGGTGGTACATGGTCCATCACCGACT 160
QY	400	CAACCAATCGCTTGTGTTTCCCTGAGTGTCTCAATGGTTAAATTAATAATCACTTAATG 459
DB	161	CCCAAAATGCTCCACTTTTCCACAGTGTGTGAATGGTTATGGATAATGAAATCAAG 220
QY	460	ATGCTTCATGGGTCTTGTATATCACTCAATCAATCAATCAATCAATCAATCAATCAAT 519
DB	221	ATGATGTTGGGGAATTGATA-----ACTATGAGGATGAATGTGTAAAGAGATG 271
QY	520	CTCTATCTTCAACATTAGCATGTATGTTGCATTAATAAGATGGAATGTTGGGGAAGATC 579
DB	272	TGTTATGATGTAGAGTGGGTAGTATGATAGGGTTAAGAGTGGGGAATTTGGTGAAGAG 331
QY	580	AAATAAATAAGCTCTAAGTTTAAATGAGTCAAACTCTGCTTCAGCTACTGAAAAAGTC 639
DB	332	AAATAAAGAAGGTATGAGGTTTATTCAGTGAATTTGCGATTTAGTGAATGATTTTTTTA 391
QY	640	AACATCTCCATTTGTTTGCATCATATTTCTGCTTCTGTTGTTGTTGTTGTTGTTGTTG 699
DB	392	TAGAGAAAGAGAGAGGTTTGTATTTATTTTGGTGGGATGATTAATTTGTTAGATTT 451
QY	700	TGGACATAAACCCTCTTTCAAAAACAAACAGATTTTATGTTTGTCTACATAAGAGGAAT 759
DB	452	TGATATGAGATTAATTTGGCGTGAAGATGCTGGATGAATGATAAGAAAAAGAGAT 511
QY	760	TGGAGCAAAAAGATGCCATTCAAATGAGATG-----GATGATCTTGGCGTAT 809
DB	512	GTGGATATTAATGTTGATAGTGAAGATTTTGAAGGGAGAGAGAGATATGTTGGGATAT 571
QY	810	ATCTCTGAAGGACTCGGTAAATTTATGTTGGAATTTGTTGGAATTTGTTGGAATTTGTTG 869
DB	572	GTTTGAAGGGGAAAAAGAAAAATAAAGATTTGGATTTTGAATTAATTAATTAAGAGAAA 631
QY	870	AATGGTTCTGTTTCAACTCACCATCAGCAACAGCTGCTGCTTTCATTAATCATCAAAAT 929
DB	632	AATGGTAAATGTTGATTTATGAGGTAAAGAGAGATGTTTAAATGATTTGGGAT 691
QY	930	CTGGTTGCTTAAATTTAAATTTCACTTTTGGACAGTTTGGTAAATGAGTCCCAACA 989
DB	692	GAAGGTTGTAAGTTATATATGTTATGTTTGTGTTTGAAGATTTAGAGGGTGAAGTTGATTTA 751
QY	990	GTTTATCTCATGATTTATTTATCCGACTTTTCTATGTTGACAAATTTGAAGATTTAGGA 1049
DB	752	GTTTATATATTTGATGAATTAAGTGGTTAGTAAATTTGTA-TTTTGAAGATTTAGGA 810
QY	1050	ATTTTCAACCATTTTCAGAGTGGAAATTTAAATAATTTTATAGTGAACATACAGATTTGG 1109
DB	811	ATTGATAGAGATTTTAAAGTGAATTTAAAGATATTTGGATGAAA-ATATAGATATTGG 869
QY	1110	GTGAACAGATGAGCAATATTCATGAGATGTTGTAACATGCTGCTTTAGCCTTTTCGGTTA 1169

```

Db      870 GTTTGCGGGA--AGAAATATGTTTGGATTGGGAATTTGGTTTGGCTTTTGATTA 927
Qy      1170 TTAAGGATCAATGGTATGAAGTTTCCCGAGATCCATTCGGCTGGAATTAATCAATGAATTA 1229
Db      928 TTTGTTGCTGATGGATATGATGTGTGTTTAAGATAAGATAAAAGATTTTGCAGGATTTG 987
Qy      1230 GCTT 1233
Db      988 GTTT 991

RESULT 12
LOCUS   CF037118/c
DEFINITION OCG40h09.yg QCG Zea mays cDNA clone QCG40h09, mRNA sequence.
ACCESSION CF037118
VERSION   CF037118.1 GI:32932306
KEYWORDS EST.
SOURCE   Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 543)
AUTHORS   Genoplante.
TITLE     Genoplante, a major partnership french program in plant genomics
JOURNAL   Unpublished (2003)
COMMENT   Contact: Genoplante
          Genoplante
          93, rue Henri Rochefort 91025 EVRY CEDEX France
          Tel: 33 1 69 47 54 00
          Fax: 33 1 69 47 54 10
          This sequence has been generated in the framework of the french
          plant genomics programme 'Genoplante' (http://www.genoplante.com)
          and http://genoplante-info.infobiogen.fr.
FEATURES             source
   1..543
    /organism="Zea mays"
    /mol_type="mRNA"
    /cultivar="F2"
    /db_xref="taxon:4577"
    /clone="QCG40h09"
    /tissue_type="embryo"
    /clone_lib="QCG"

ORIGIN
Query Match      7.0%; Score 195.8; DB 6; Length 543;
Best Local Similarity 61.1%; Pred. No. 2.4e-34;
Matches 317; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

Qy      1505 CTCATCAATATTAGTAACACTGATTACCTAAGGTGGCTCTTGAAGATTTCTACACCTG 1564
Db      538 CTCATCTTTTCGATCAATCAAGAAATCTCTAGCTTGGCATCGAAGATTTCACTTCTC 479
Qy      1565 CCAATCTATTATTATCGTGAAGAAATTAAGAGCTTGAAGGTGGGTGGAGAGATAAGTT 1624
Db      478 TCACAGGTGTTACGGGATGAACCTCGCATCTTGATAGTTGGGTGAAGAGACAACT 419
Qy      1625 GGACCACTCAAGTTTGTCTAGGCAAAAGACCGCTACTGTTATTCTCTGTTGCTGCAAC 1684
Db      418 GGACCACTCAAAATTTGCTGGCAGAAACTGACATATTGCTATCTGCTGCTGCTAC 359
Qy      1685 ACTTTCGTCCTCCGAATATCAGATGCGGTATTTTCATGGGCCAAAATGSCATATTAAC 1744
Db      358 CGTATTTCCTCTCAATTTGCTGACGCTCGCATTTTCATGGGCCAAAATGCTGCTCAC 299
Qy      1745 TACAGTAGTTGATGACTTTTTTGTATATCGGTGGTACAAATGATGATGATGATGATGATGAT 1804
Db      298 AACTGTGTTGATGACTTCTTCGATGTTGGTGGATCAAAAGAGAAATGAAAACCTGAT 239
Qy      1805 TCAATGTTGAAAAATGGAATGTAGATGTGCAAGAGATTTGTTTCAGAGCATGTTG 1864

```

```

Db      238 AGCACTAGTTGAGAAATGGCATGCCACCATGCAGTTCTAGTTCTATTGGAACAGTGAA 179
Qy      1865 GATTTTATTTTAGCAATTAAGAGTCAATCTGTGGATTGGAGATGAAGCTTTTAAATG 1924
Db      178 AATAGTATTTTCTGTTATTAACAACAGTGAACCATCTTGAGCAATGGCTTCTGCGAC 119
Qy      1925 GCAAGCGCGCATGTAACTAGCCATGTTATTCAACACTTGGTTGGAACTAATGAATAGTAT 1984
Db      118 ACAAGGCCGTGATCTTACAACCCACTAGTAGAAATATGGCTGGATTTGTTAAGATCTAT 59
Qy      1985 GTTGAGAGAAGCTATATGGAACAAGAGATGCTTATGTGCC 2023
Db      58 GATGTCGAGGAGAGATGGCAGAGATGCCAATATGTACC 20

RESULT 13
LOCUS   CF197159
DEFINITION Ir64vii.5.E05 IR64-VI Oryza sativa (indica cultivar-group) cDNA clone IR64-VII.5E05 5', mRNA sequence.
ACCESSION CF197159
VERSION   CF197159.1 GI:33381816
KEYWORDS EST.
SOURCE   Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 619)
AUTHORS   Ventelon-Debout,M., Thi-Thu Huang,N., Wissocq,A., Berger,C.,
          Michele,L., Piegu,B., Cooke,R., Ghesquiere,A., Delsey,M. and
          Brugidou,C.
TITLE     Metabolism and Photosynthesis gene expression is affected in
          response to rice yellow mottle virus infection in Oryza sativa
          indica and Japonica cultivars
JOURNAL   Mol. Genet. Genomics (2003) In press
COMMENT   Contact: Brugidou C
          IRD
          BP 64501, 34394 Montpellier cedex 5 France
          Tel: 33 4 67 41 67 39
          Fax: 33 4 67 41 61 81
          PCR Primers
          FORWARD: 5'AATTAAACCTCCTCAATAAGGG3'
          BACKWARD: 5'CGGATATCACTCAGCATATG3'
          Plate: 1.5 row: E column: 5.
FEATURES             source
   1..619
    /organism="Oryza sativa (indica cultivar-group)"
    /mol_type="mRNA"
    /cultivar="IR64"
    /db_xref="taxon:39946"
    /clone="IR64-VII.5E05"
    /tissue_type="leaves"
    /dev_stage="3 weeks-old plants"
    /lab_host="SOLR strain (Stratagene)"
    /clone_lib="IR64-VI"
    /note="Vector: pBluescript SK (+/-); Site 1: Eco RI;
          Site 2: Xho I; Sample name: IR64-VI; Plant growth place:
          greenhouse; Soil conditions: standard; Sowing date:
          28/02/2000; Harvesting date: 17/03/2000; Stress date:
          15/03/2000; Cultivar: ir64; Description: IR64-RYM
          inoculated leaves cDNA library; experimental condition:
          rymv inoculation;"

ORIGIN
Query Match      6.8%; Score 190.4; DB 6; Length 619;
Best Local Similarity 56.1%; Pred. No. 4.4e-33;
Matches 345; Conservative 0; Mismatches 267; Indels 3; Gaps 1;

Qy      1393 TGAATAATGCTCTTAAGTTCCCTATCAATACCGGTTTAGAACGCATAAACACTAGACGA 1452
Db      2 TGAAGTATGCTCTTAAGTTCCCTTCTACACCACCTGGACCGCTCTAGATCATCAGAGGA 61

```


/db_xref="taxon:4577"
/clone="QCG9g06"
/tissue_type="embryo"
/clone_lib="QCG"

ORIGIN

```
Query Match      6.6%; Score 185; DB 6; Length 515;
Best Local Similarity 61.4%; Pred. No. 7.6e-32;
Matches 315; Conservative 0; Mismatches 195; Indels 3; Gaps 1;

Qy 1380 ATTCAACAAGAGGTGGAAAAAGCTCTTAAGTTCCCTATCAATACCGGTTTAGAACGCATA 1439
Db 5 ATACTTGAAGAGGTAGAGTATGCCCTTAAGTTTCCCATTTATGCCGTAGTAGATCCCTCTC 64

Qy 1440 AACACTAGACGAATATACAGCTTTTACAAATGTAGACAATACAAGAATTCGAAACTACA 1499
Db 65 AACACAGAGAAGATGGAGCAATTTGGATGTAGGGGTTCTCAGACGCTAAGACAAAA 124

Qy 1500 TATCACTCATCAAAATATTAGTAACACTGATTACCTTAAGTTGGCTGTTGAAGATTTCCTAC 1559
Db 125 TATCTGCCATGTCAATGAAGTCA---AGACATCCTTTGGCTGTTGAAGATTTCCTGT 181

Qy 1560 ACCTCCCAATCTATTATTCGTGAAGAATTAAGGTCTTGAAGGTGGGTGGTAGAGAAAT 1619
Db 182 TTTTCTCAATCTATTACAGGACGAACTACAGAAACATCATTAGTTGGGAGAAAGAAAT 241

Qy 1620 AAGTTGGACCAAGCTCAAGTTTGTAGGCAAAAGACCGCTACTGTATTCTCTGTTGCT 1679
Db 242 AGGATGGACCAAGCTACATTTTGTGGGCAAAAGGCTGGCAATATGCTATCTCGTGTGCT 301

Qy 1680 GCAACACTTTCGTCTCCGAATATCAGATGGCGTATTTTCATGGGCCAAAAATGGCATA 1739
Db 302 ACCACCATATCCCTCATGAATTTGTCTGATGCTGCGTTGCATGTGCCAAAGCATCATG 361

Qy 1740 TTAACACAGTACTGATGATGACTTTTGTGATTCGGTGGTACAATCGATGAATTCACCAAC 1799
Db 362 CTCACGGTTGTAGTTGATGACTTCTTCGATGTTGGTGGATCAAAAGAGAAACAAGAAAT 421

Qy 1800 CTGATTCAATGTGTGAAAAATGGAATGTAGATGTGCAAGAGGATGTTGTTTCAGAGCAT 1859
Db 422 CTCATCGAATTAGTTGAGAACTGGGATGAGCACCAAAAGTTGAGTTCTGTTCGGAGAA 481

Qy 1860 GTTCGGATTTTATTATTAGCAATTAAGATGCA 1892
Db 482 GTAGAAATAGTTTCTATGCTGTCTATAATACA 514
```

Search completed: October 17, 2004, 03:24:33
Job time : 8685.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 23:38:29 ; Search time 1297.73 Seconds
(without alignments)
11005.915 Million cell updates/sec

Title: US-10-041-018-361

Perfect score: 2792

Sequence: 1 cccactcatctttatcaac.....aaaaaaaaaaaaaaaaaaaaa 2792

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3403857 seqs, 2557783690 residues

Total number of hits satisfying chosen parameters: 6807714

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2792	100.0	2792	16	US-10-041-018-361	Sequence 361, App
2	2582.6	92.5	3117	16	US-10-041-018-362	Sequence 362, App
3	1428.6	51.2	2594	16	US-10-041-018-368	Sequence 368, App
4	809.4	29.0	2638	16	US-10-041-018-367	Sequence 367, App
5	782.2	28.0	2658	16	US-10-041-018-378	Sequence 378, App
6	782.2	28.0	2658	16	US-10-041-018-381	Sequence 381, App
7	732.4	26.2	2506	16	US-10-041-018-377	Sequence 377, App
8	561.2	20.1	2554	16	US-10-425-114-3441	Sequence 3441, App
9	433.2	15.5	2085	16	US-10-425-114-17004	Sequence 17004, App
10	392.4	14.1	2029	16	US-10-425-114-5986	Sequence 5986, App
11	379.4	13.6	2223	16	US-10-259-194A-107	Sequence 107, App
12	356.4	12.8	2070	17	US-10-437-963-19081	Sequence 19081, App
13	330	11.8	1559	16	US-10-041-018-375	Sequence 375, App

14	315.6	11.3	3666	17	US-10-437-963-72152	Sequence 72152, App
15	288.8	10.3	1150	16	US-10-425-114-729	Sequence 729, App
16	276.8	9.9	1008	16	US-10-425-114-2257	Sequence 2257, App
17	275	9.8	1615	17	US-10-437-963-94395	Sequence 94395, App
18	264.8	9.5	2193	17	US-10-437-963-72156	Sequence 72156, App
19	241.6	8.7	2313	17	US-10-437-963-36043	Sequence 36043, App
20	239.4	8.6	929	16	US-10-424-599-51115	Sequence 51115, App
21	238.4	8.2	1157	17	US-10-437-963-19703	Sequence 19703, App
22	198.4	7.1	711	17	US-10-437-963-72153	Sequence 72153, App
23	194.8	7.0	2861	9	US-09-895-752-55	Sequence 55, App
24	194.8	7.0	2861	9	US-09-887-586A-55	Sequence 55, App
25	194.8	7.0	2861	9	US-09-903-012-55	Sequence 55, App
26	194.8	7.0	2861	10	US-09-900-797-55	Sequence 55, App
27	194.8	7.0	2861	11	US-09-893-820-55	Sequence 55, App
28	194.8	7.0	2861	13	US-10-041-007-3	Sequence 3, App
29	194.8	7.0	2861	16	US-10-041-018-365	Sequence 365, App
30	172.8	6.2	439	16	US-10-041-018-373	Sequence 373, App
31	172	6.2	577	16	US-10-259-194A-491	Sequence 491, App
32	155.4	5.6	575	17	US-10-767-701-30695	Sequence 30695, App
33	153.8	5.5	2700	9	US-09-895-752-43	Sequence 43, App
34	153.8	5.5	2700	9	US-09-887-586A-43	Sequence 43, App
35	153.8	5.5	2700	9	US-09-903-012-43	Sequence 43, App
36	153.8	5.5	2700	10	US-09-900-797-43	Sequence 43, App
37	153.8	5.5	2700	11	US-09-893-820-43	Sequence 43, App
38	153.8	5.5	2700	16	US-10-041-018-363	Sequence 363, App
39	133.8	4.8	2241	13	US-10-041-007-38	Sequence 38, App
40	133.8	4.8	2388	13	US-10-041-007-36	Sequence 36, App
41	133.8	4.8	2445	13	US-10-041-007-34	Sequence 34, App
42	133.8	4.8	2622	13	US-10-041-007-32	Sequence 32, App
43	133.8	4.8	2705	13	US-10-041-007-1	Sequence 1, App
44	133.8	4.8	2705	16	US-10-041-018-397	Sequence 397, App
45	130	4.7	339	11	US-09-955-216-20	Sequence 20, App

ALIGNMENTS

RESULT 1
US-10-041-018-361
; Sequence 361, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Biterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US/10/041.018
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 361
; LENGTH: 2792
; TYPE: DNA
; ORGANISM: Stevia rebaudiana
US-10-041-018-361

Query Match	100.0%;	Score 2792;	DB 16;	Length 2792;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2792;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	CCCACTCATCTTTATCAACAAATACCATCGTTCTGCCACCGGAAGACTGATAGCGACG	60	
Db	1	CCCACTCATCTTTATCAACAAATACCATCGTTCTGCCACCGGAAGACTGATAGCGACG	60	
Qy	61	GACCTGACGACGCTTTAACTCTGTGGACCACTTAACAAAATCCGACCATCAGATCT	120	
Db	61	GACCTGACGACGCTTTAACTCTGTGGACCACTTAACAAAATCCGACCATCAGATCT	120	
Qy	121	ACTCCGGTGACAGTTTTCATTTGCAATTTAGTAAATCAGCAAACTAAACATCAATGATC	180	
Db	121	ACTCCGGTGACAGTTTTCATTTGCAATTTAGTAAATCAGCAAACTAAACATCAATGATC	180	

Db 1786 GCTTACAAATGAGCAATACAAAGAACTCTGAAAACATCATATCACTCATCAAAATTAG 1845
Qy 1520 TAACTACTGATTAACCTAAGGTTGGCTGTGGAAGATTTTACACCTGCAATCTATTATCG 1579
Db 1846 TAACTACTTATTACCTAAGGTTGGCTGTGGAAGATTTTACACCTGCAATCTATTATCG 1905
Qy 1580 TGAAGAAATTAAGGTTCTTGAAGGTTGGCTGTGGAAGATTTTACACCTGCAATCTATTATCG 1639
Db 1906 TGAAGAAATTAAGGTTCTTGAAGGTTGGCTGTGGAAGATTTTACACCTGCAATCTATTATCG 1965
Qy 1640 TCTTAGGCAAAAGACCGCTACTGTTATTTCTCTGTTGCTGCAAAACATTTCTGCTCCCGA 1699
Db 1966 TCTTAGGCAAAAGACCGCTACTGTTATTTCTCTGTTGCTGCAAAACATTTCTGCTCCCGA 2025
Qy 1700 ATTATCAGATGCGGTATTTCTATGCGCCAAATAATGGCAATTAATACAGTAGTGTGATGA 1759
Db 2026 ATTATCAGATGCGGTATTTCTATGCGCCAAATAATGGCAATTAATACAGTAGTGTGATGA 2085
Qy 1760 CTTTCTTGTATGATCGGTGTACATCGATGAATTCACCAACCTGATTCATATGTTGAAAA 1819
Db 2086 CTTTCTTGTATGATCGGTGTACATCGATGAATTCACCAACCTGATTCATATGTTGAAAA 2145
Qy 1820 ATGGAATGTAGATGTCGACAAAGGATTTGTTTCAGAGCATGTTTCGGATTTATTTTACG 1879
Db 2146 ATGGAATGTAGATGTCGACAAAGGATTTGTTTCAGAGCATGTTTCGGATTTATTTTACG 2205
Qy 1880 ATTAAGAATGCAATCTGTTGGATGAGATGAAGCTTTTAATGCGAAGCGCGCATGT 1939
Db 2206 ATTAAGAATGCAATCTGTTGGATGAGATGAAGCTTTTAATGCGAAGCGCGCATGT 2265
Qy 1940 AACTAGCCATGTTATTCAAACTGTTGGATGGAATGGAATGGAATGGAATGGAATGGAAT 1999
Db 2266 AACTAGCCATGTTATTCAAACTGTTGGATGGAATGGAATGGAATGGAATGGAATGGAAT 2325
Qy 2000 ATGCAACAGAGATGCTTATGTCGCAACATTAATGAATATATGGAACAGCTTTACGTGC 2059
Db 2326 ATGCAACAGAGATGCTTATGTCGCAACATTAATGAATATATGGAACAGCTTTACGTGC 2385
Qy 2060 ATTTGCAATGAGCCGATGTTCAAGCGGCTATTTTACTTTTGGGGGCCCAATATCAGA 2119
Db 2386 ATTTGCAATGAGCCGATGTTCAAGCGGCTATTTTACTTTTGGGGGCCCAATATCAGA 2445
Qy 2120 GAGATGTTGGAAGCTCTGAATATCATATCATATCTATTAATGCTAATGAGCAGCGGTGC 2179
Db 2446 GAGATGTTGGAAGCTCTGAATATCATATCATATCTATTAATGCTAATGAGCAGCGGTGC 2505
Qy 2180 ACTTCTAAACGATATCCATAGCTTCAAGAGGGAATTTAAGGAAGGCAAAATTAACCGCGT 2239
Db 2506 ACTTCTAAACGATATCCATAGCTTCAAGAGGGAATTTAAGGAAGGCAAAATTAACCGCGT 2565
Qy 2240 AGCATTTGCAATTCAGATGACGAGAAAGTGGGAAGTGGGAAGAGAGTGTGAGGAGAT 2299
Db 2566 AGCATTTGCAATTCAGATGACGAGAAAGTGGGAAGTGGGAAGAGAGTGTGAGGAGAT 2625
Qy 2300 GATGATGATCATTTAAAGAAAGAGAAATTAATGAAATTAATTTTCAAGAAATAGG 2359
Db 2626 GATGATGATCATTTAAAGAAAGAGAAATTAATGAAATTAATTTTCAAGAAATAGG 2685
Qy 2360 TAGCATTTGTTCTTAGAGCTTTGTAAGATGATTTTGGAAATGATGTCACGTTGTAATTT 2419
Db 2686 TAGCATTTGTTCTTAGAGCTTTGTAAGATGATTTTGGAAATGATGTCACGTTGTAATTT 2745
Qy 2420 TTTTTCGCAAAACGATCAGCGTTTACTGGAACACGATCTTCTGATCTGTGAGGAGCAT 2479
Db 2746 TTTTTCGCAAAACGATCAGCGTTTACTGGAACACGATCTTCTGATCTGTGAGGAGCAT 2805
Qy 2480 CATTTACAAACCGTTGTTGCTGTGGAATGAAATGAAGAAACAAAGGTAATTTGCGCTTAA 2539
Db 2806 CATTTACAAACCGTTGTTGCTGTGGAATGAAATGAAGAAACAAAGGTAATTTGCGCTTAA 2865
Qy 2540 AAGATGATTAATAATATGCTCTTGAACGCGGCTGTTGATGTTGATTAATAAGGTT 2599

Db 2866 AAGATGATTAATAATATGCTCTTTGACGGGGTACTGTTGATGTTGATTAATAAGGTT 2925
Qy 2600 GTAGTTGTATATAAAGGTAATAGCTAATCAATAGAAAGCTTAAGTTAATTAAGTTTCC 2659
Db 2926 GTAGTTGTATATAAAGGTAATAGCTAATCAATAGAAAGCTTAAGTTAATTAAGTTTCC 2985
Qy 2660 CTCGTGTCCACACACGCGTGAGGTTCTTGTAAACGAGTTTATTTATGTTTGTAGTACAC 2719
Db 2986 CTCGTGTCCACACACGCGTGAGGTTCTTGTAAACGAGTTTATTTATGTTTGTAGTACAC 3045
Qy 2720 TTGGTGGTTGTTGCTGTTTATCTCTAAGAGGCTTTCACCTTGTCTTAAAAAAA 2776
Db 3046 TTGGTGGTTGTTGCTGTTTATCTCTAAGAGGCTTTCACCTTGTCTTAAAAAAGAA 3102

RESULT 3
US-10-041-018-368
; Sequence 368, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080051/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; PRIOR FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 368
; LENGTH: 2594
; TYPE: DNA
; ORGANISM: Lactuca sativa
US-10-041-018-368

Query Match 51.2%; Score 1428.6; DB 16; Length 2594;
Best Local Similarity 75.8%; Pred. No. 0;
Matches 1833; Conservative 0; Mismatches 529; Indels 57; Gaps 3;
Qy 238 CAGCTATTATCATACAGCATCAATTCACATGTCGAGCAAACTAATCCACATTAATCTGATCA 297
Db 187 CATCGGGTCTTCTGTCAGCTTCTTCAAGCTGCAAGTAAATCTTCTGTCATGACCC 246
Qy 298 TTGATCAACCAAGAAACGATCCAAAAACAGTTTAAAAATGAGAAATTTCTGTTCTT 357
Db 247 TTGATGTGACCAAGAAACGATCCGAAAGCTGTTCAACATGTGGAAGTTTCTGTTCTT 306
Qy 358 CATATGACACAGCATGGGTAGCCATGGTCCCTTCTCCAAACTCACCACATCGCCTGTT 417
Db 307 CATATGACACAGCTTGGGTAGCCATGGTCCCTTCTCCAAACTCACCACATCGCCTGTT 366
Qy 418 TCCTCGAGTCTCAATGCTTAAATTAATCAAGCTTAATGATGTTTCATGGGCTCTG 477
Db 367 TCCTCGAGTCTCAATGCTTAAATTAATCAAGCTTAATGATGTTTCATGGGCTCT 424
Qy 478 TTAATCACATCATATATCAATCAATCACCCTGCTTAAAGATTTCTATCTTCAACATTAG 537
Db 425 -----TCCTCGATCAGTCCCAATTAATTAAGATATCTCTCTTCAACATTAG 474
Qy 538 CATGTATGTTGATTAATAAGATGATGTTGGGAGATCAATTAATAAAGGCTCA 597
Db 475 CATGTGATGTTGATTAATAAGATGATGTTGGGAGATCAATTAATAAAGGCTCA 534
Qy 598 GTTTTATGAGTCAAACTTGTGCTTCACTACTGAAAAAGTCAACCTCTCCATTTGTT 657
Db 535 ATTACATTGAGTCAAACTTGTGCTTCACTACTGAAAAAGTCAACCTCTCCATTTGTT 594
Qy 658 TTGACATCATATTTCTGTTGTTGTTGAGTATGCGAAAAAATTGGACATTAACCTCTTT 717
Db 595 TTGACATCATATTTCTGTTGTTGTTGAGTATGCGAAAAAATTGGATATTAACCTCTTT 654
Qy 718 CAAACAAACAGATTTTATGTTGATGCTACATAAGAGGGAATTCGAGCAAAAAAGATGCC 777

; SOFTWARE: PatentIn version 3.1									
; SEQ ID NO 367									
; LENGTH: 2638									
; TYPE: DNA									
; ORGANISM: Cucumis sativus									
US-10-041-018-367									
Query Match 29.0%; Score 809.4; DB 16; Length 2638;									
Best Local Similarity 62.0%; Pred. No. 6.7e-180;									
Matches 1410; Conservative 0; Mismatches 811; Indels 54; Gaps 6;									
QY	267	GGTGACAACTAAATCCCACTAAATCTGATCAATGTATACAAACAAAGAAAGCGATCCAAAAA	326						
DB	159	GGCACAATAAATACTGAGCTTTGGCTTTGAGAAACAAAGAAAGATATAAAAA	218						
QY	327	CAGTTTAAATGTGAAATTTCTGTTCTTCAATATGACACAGCATGGGTAGCATGGTC	386						
DB	219	TTGTTCAAAATGTGAACTTTCAATTTCTGCATATGATCTGCATGGGTGGCAATGGTC	278						
QY	387	CCTTCTCCAAATCCACCAATCGCTTGTTCCTTGAGTGTCTCAATTTGGTTAAATTAAT	446						
DB	279	CTTCTCCAAATCTCTTAATAAACCCTTTTTCTTGAGTGATTAACCTGGGTATTAGAT	338						
QY	447	AATCAGCTTTAATGATGGTTTCATGGGTCTTGTTAATCACATCATATAATCATACCCG	506						
DB	339	CATCAAAACCTTGATGGGTCAATGGG-----CATACTCCATGACCATCAGTTG	386						
QY	507	TTGCTTAAAGATCTCTATCTTCAACATGATGATGATTTGTCATTAATAAGATGAAT	566						
DB	387	GTGATGAAGCCACTCTCTATCCACATTAGCATGTGTTCTTACTCTTAAGCGATGGAT	446						
QY	567	GTTGGGAGAGATCAATAATAAGCTTAAGTTTTATTGAGTCAAAATCTTGCTTCAGCT	626						
DB	447	ATCGGTGATGATCATATGAGCAAGGCCCTTAGTTTTATCAAGTCTTAATATAGCTTCAGCT	506						
QY	627	ACTGAAAAAGTCAACCATCTCCATTTGGTTTGTGACATCATATTTCCCTGGTTTGTGAG	686						
DB	507	ACTGATGAGAACTCAACGTTCTCTGTTGGATTTGACATAATTTTCCCTGGTATGATTGAG	566						
QY	687	TATGCAAAAACCTTGACATTAACCTCTTCAACCAACACAGATTTTATGTTGATGCTA	746						
DB	567	TATGCTAAAGACTGAATTTGATTTGATCTACCTTTGGCATCAATGAATGGATGCTTTGGTT	626						
QY	747	CATAAGAGGAAATGAGCAAAAAGAT-----GCCATTCAAATCAGATGGATGATAC	800						
DB	627	CAAAAGAAAGTTGAGCTTTAGAGCTGCTGTAGCACTCTGAGAGAGAAAGCCTAT	686						
QY	801	TTGGCGTATATCTCTGAAGGACTCGGTAATTTATATGATTTGAATATGTTGAAGAAATAT	860						
DB	687	TTAGCGTATGTTTCAGAAAGAAATGGAATAATTACAGGACTGGGAATGGTCATGCGATAT	746						
QY	861	CAGATGAATAATGGTTCTGTTTCAACTCACCATCAGCAACAGCTGCTGCTTTCAATTAAT	920						
DB	747	CAAGGAAGAACCGATCACTGTTTATGTTTCCATCCACCACCGCAGTGGCTTTATGCAAC	806						
QY	921	CATCAAAATCCTGGTTGCTTAATTAATTTAAATTTCACTTTTGGCAAGTTTGGTAATGCA	980						
DB	807	AGAAATGATGAGCTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	866						
QY	981	GTCCCAACAGTTTATCCTCATGATTTATTTATTCGCACTTTCTATGTTGACACAAATGAA	1040						
DB	867	GTTCGCCCAATATATCTCTTGATATATATGCTGCTTTTACATGTTGATAGCCCTTCAA	926						
QY	1041	AGATTAGGAATTTCAACCAATTTGAGTGGAAATTAATAATTTTATGATGAAACATAC	1100						
DB	927	AAATCGGGGATGATGGCCATTTCAAGATGAGATGAGATGAGATGATGATGATGATGATGAT	986						
QY	1101	AGATGTTGGGTGAAACGAGATGAGCAAAATATTCAATGAGTGTGTTAATCATGTGCTTTAGCC	1160						
DB	987	AGCTGTTGATGCAAGGGAGGAAACATATTCCTAGATGCTTCAACTTTGTGCAATGGCC	1046						
QY	1161	TTTCGGTTATTAAGGATCAATGGGTATGAAAGTTTCCCGATCCATGGCTGGAATTAAT	1220						

DB	1047	TTCCGAGTTTACGTTTGAAGGATATGATGTTTCTTCAGATCAATGACTCAATCTCA	1106						
QY	1221	AA-----TGAATTAGCTTTGAAGACGAATATGACCTCTGAA	1259						
DB	1107	GAGGTCTCTTTTCAAAATGCTCGAGGACATTTAAAGACTTTAGTGCCTCATTAGAG	1166						
QY	1260	ACATATCATGCTGCACA---TATATATACCAAGAGGATTTATCTTCTGAAAAAATAATC	1316						
DB	1167	TTATTTAAGGCTCCAGATTTATCAATTTATCCGATGAGTTTATCTCGAAAAATAAATC	1226						
QY	1317	TTGAGTCAAGCTGATTTCTCTCAAGAGATATATCCACTG-----ATTCAACAGG	1367						
DB	1227	CTTGGACTAGTCTGTTCTCTGAATCATGATTTATCTAGTGTCTAGTTCAATCTGATAGA	1286						
QY	1368	CTTTCTAAATTAATTTCAAAAGAGGTGAAATGCTTTAAGTTCCTTCAATCAATCCGCT	1427						
DB	1287	ACTGAGAGACTCGTGAACACAGAGGAGTTAATGTTTCGAGTTCCCTTATATTAATCACT	1346						
QY	1428	TTAGAACGCATAAAACATAGACGAATATACAGCTTTTACAAATGTATAGACAAATCAGAAAT	1487						
DB	1347	CTGGAACGCTTATCAAAATAAGCGACACTGGAAAGTTACAGTGGAGACATTTGAGGAT	1406						
QY	1488	CTGAAAACTACATATCATCTCAATATATTAGTAACACTGATTACTTAAGTTGCTGTT	1547						
DB	1407	TCAAAAACAGCATATGCTGCTTAAATTTTGTCTATCAAGATTTCTTGGAACTTGCTGA	1466						
QY	1548	GAGATTTCTACACCTGCCAATCTATTTATCTGGAAGATTTAAAGTCTTCAAAAGTGG	1607						
DB	1467	GAGATTTCTCAATCCCTGCAAGGCTATCATCCAGAGAACTGAAAGAGCTTGAAAAATGG	1526						
QY	1608	GTGTAGAGAAATPAAGTTGGACCAAGCTCAAGTTTGTCTAGCAAAAAGACCCCTCTGTTAT	1667						
DB	1527	GTTCATGAAAAACAAATTTGCAAAAGTTGCAAGACAGAAAGTTAGCTGACTGCTAT	1586						
QY	1668	TTCTCTGTGTCGCAACACTTTCTGCTCCGAAATTTACAGATGCGGATTTTCATGGGCC	1727						
DB	1587	TTTTCTGTGTCGCAACCTTACCTCCAGAACTTTGTGATGCGCGCTTATCATGGGCA	1646						
QY	1728	AAAAATGGCATTTAATACAGTAGTTGATGACTTTTGTGATATCGGTGGTCAATCGAT	1787						
DB	1647	AAAAATGGGTACTCAACCTGTTGTGATGATTTCTTGTGATGTTGGAGAACTGGAAG	1706						
QY	1788	GAAATGACCAACTGATTCATTTGTGAAAAATGGAATGTAGATGCAAGAGATTTGT	1847						
DB	1707	GAAATGGTAAACCTTATATACAATTTGTGGAAGAAAGTGGATGCGAGTGGGAAACGGGTAC	1766						
QY	1848	TGTTTCAGACATGTTCCGATTTTATTTTAGCATTAAGATGCAATCTGTTGGATTGA	1907						
DB	1767	TGTTCCAGGAGTTGAGATTTATTTCTTTCACATTCATAGCACAATTTGTGAAATAGGA	1826						
QY	1908	GATGAAGCTTTTAAATGGCAAGCGCGCATGTAACTTAGCATGTTTATTCAAACTTGGTTG	1967						
DB	1827	AAAAAGCTTTTACCTTGGCAAGGACGCGCTGATGAGGAAATGTTATCGATATTTGGTTG	1886						
QY	1968	GAACTAATGAATGATGATTTGAGAGAGCTATATGAGCAAGAGATGCTTATGTGCCAACA	2027						
DB	1887	GCTTTGCTCGACTCAATGAGGAAAGAGCTGAAATGTTGAAAAATAAGGTAGTGCATCA	1946						
QY	2028	TTAATGAATATATGAAAAAGCTTACCTGCTCATTTGCTATTTAGCCCGGATTTGCAAGCGG	2087						
DB	1947	TTGGATGATGATCATGAAAAATGGCTATGATTTATTTGCTTTTGGGACCTATAGTCTTCCA	2006						
QY	2088	GCTATTATCTTTTGGGGGCCAAATTTATCAGAGGAGATTTGTTGAAAGCTCTGAATATCAT	2147						
DB	2007	AGCTCTACTTTTGTGACCTTAAGCTTCCAGAGGAAATTTGTTGAAATTTGTAATACAAG	2066						
QY	2148	AATCTAATTAAGCTAATGAGCGCGAGGTGCACTTTCTTAAACGATATCATAGTCTCAAG	2207						
DB	2067	AAGCTCTTTAAGCTGATGAGCACTTCTGCGCGCTTCTGAATGATATCTGAACTTTTGTAT	2126						
QY	2208	AGGGAATTTAAGGAAGGCAAAATTTAAACGCGGTAGCATTTGATTTAGTAACGGAGAAAGT	2267						
DB	2127	AGAGTCCAGGAGGGAATTAATGCTTCTCTATACATGATTAGTCCGCGGTGTT	2186						

1664 GGCAGGACCCCTTATCGATCCTGAACTTCATGCTCGCATAGCATGGGCACAAAATGG 1723
1736 CATATTAACACTACGATGATGATGATCTTTTGGATATCGGTGTACATGCAATGATGAC 1795
1724 TGTGCTCAGACCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1783
1796 CAACCTGATCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1855
1784 TAACTCTTATGAAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1843
1856 GCAATGTTGGAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1915
1844 GACGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1903
1916 TTTTAAATGGGAGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1975
1904 TTTAGTATGGGAGGACGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1963
1976 GAATAGTATGTTGAGAGAGCTATATGAGACAGATGCTTATGTCACCAACATTAATGA 2035
1964 GAAGGTGATGAGAAAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2023
2036 ATATATGAAAACGCTTACGTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 2095
2024 ATATATGAAAACGCTTACGTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 2083
2096 CTTTGTGGGCGCAAT 2155
2084 CTTTGTGGGCGCAAT 2143
2156 TAAGCTAATGAGCAAGCGCTGATCTTCAACAGATATCCATGATCTTCAAGAGGGAAT 2215
2144 TAAGCTGATGAGCAAGCGCTGATCTTCAACAGATATCCATGATCTTCAAGAGGGAAT 2203
2216 TAAGGAAAGCAATTAACCGCGTAGCATTTGCAATTTGCAATTTGCAATTTGCAATTTGCA 2275
2204 CAAAGAGGAAAGCAATTAACCGCGTAGCATTTGCAATTTGCAATTTGCAATTTGCAAT 2263
2276 GGNAGAGAGGTTGCGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 2335
2264 CAAAGAGGAGGCAATTAACCGCGTAGCATTTGCAATTTGCAATTTGCAATTTGCAAT 2323
2336 GAAATTAATTTTGAAGAAATGTTAGCATTTGTTAGCATTTGTTAGCATTTGTTAGCATTT 2395
2324 GGGTTAGTTTTCAGGAGGAGCACTACAA---TTCCAGAGCTTTGTAAGGATTTGTTCTG 2380
2396 GAACATGTGTCAAGTTGTAATTTTTCAGCAACGATGACGGTTTTCAGCAACGATGACGG 2455
2381 GAAATGATGTCATTTGTAATCTATTTTACATGGAAGATGATGGGTACACTTCAAATAG 2440
2456 GATCTGTGATGTTGAGAGGAGATCTTTTACACCCGTTGGTCTTG 2502
2441 GTTGATGAACACTGTAAAGCCATGTTTGAACCAACCCATGGATCTGG 2487

RESULT 6
US-10-041-018-381
; Sequence 381, Application US/10041018
; Publication No. US2004007232A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080051/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 381
; LENGTH: 2658

TYPE: DNA
ORGANISM: Cucurbita maxima
US-10-041-018-381
Query Match 28.0%; Score 782.2; DB 16; Length 2658;
Best Local Similarity 61.3%; Pred. No. 1.7e-173; Indels 54; Gaps 6;
Matches 1390; Conservative 0; Mismatches 823;
QY 275 AACTATCCCACTAATCTGATCATTTGATACAAAGAACCAAGATCCCAAAACAGTTTAA 334
Db 236 AACTAAACTGAGCTTTGCACTTTGAAGAAACAAAGAAAGATTAAGAAATTTGTCGA 295
QY 335 AATGTAGAAAATTTCTGTTTTCATATGACACAGATGGGTAGCCATGGTCCCTTCTCC 394
Db 296 CAAGGTTGAACCTTTTCAGTTTTCATATGATAGTGGGTGGCAATGGTTCCTTCTCC 355
QY 395 AACTCACCACCAATCGCTTGTTCCTGAGTGTCTCAATTCGTTAAATTAATACAGCT 454
Db 356 AACTCTCTCAACCAACCTCTTTTCCCGAGTGATATAACTGGGTATTTAGATAGTCA 415
QY 455 TAATGATGTTTCATGGGTTCTTGTAAATCACACTCATAATCATATCAATCAACCGTTC 514
Db 416 TGTGATGCTCATGGGCTTACTCCACAAC-----GATCAGTTGCTGATGA 463
QY 515 AGATCTCTATCTTCAACATTTAGCATTTGTTGCAATTTAAAGAGATGGAATTTGGGA 574
Db 464 GGCCAACTCTTATCTACATTTAGCATTTGTTTACTCTTAAAGCGGTGGAATTTGGGA 523
QY 575 AGATCAATAATAAGGTTCTAAGTTTATTTGAGTCAATTTCTGTTTTCAGTACTGAAA 634
Db 524 TGATCATATAGCAAGGCTTTGATTTTATCAAGTCTAATATAGCTTCACTACTGATGA 583
QY 635 AAGTCAACCATCTCCCATTTGTTTGGATCATATTTCTGTTTTCGTTTTCAGTATCGAA 694
Db 584 GRACCAACGTTCTCCGTTGGATTTGACATTTATTTCCCTGGCATGATTGAGTATGCTA 643
QY 695 AACTTTGGACATAAACCTCTCTTCAAAACAAACAGATTTTGTGATGCTACATAGAG 754
Db 644 AGACTTTGAATTTGAATCTACCTTGGCACCAGCAACGTTGGATGCCCTTGGTTCGA 703
QY 755 GGAATTTGGAGCAAAA-----AGATGCCATTCAAATGAGATGGATGATACATTGGCGTA 808
Db 704 AGATTTGGAGCTGAGAGCTGAGAGCACTCTGAGGTGGAAAGCCATTTTGGGTA 763
QY 809 TATCTCTGAAGGACTCGGTAATTTATGATTTGGAATATGGTGAAGAAATATCAGATGA 868
Db 764 TGTTCAGAGGAATTTGGAAGTTACAGGACTGGGATATGTCATGCAATATCAAGGAA 823
QY 869 AATGTTCTGTTTTCACACTCAGCAACAGCTGCTGCTTTTCAATTAATCATCAAAA 928
Db 824 GAATGGATCACCTGTTTAAATTTCTCCACTACGGCAGCGGCTTTTATGCAATAGAATGA 883
QY 929 TCTGTTGTTCTTAATTTTAAATTTCACTTTTGGACAAGTTTGGTAATCCAGTCCCAAC 988
Db 884 TGATGGCTGTTTGGATTTCTCGCTACCTCTTACAAAGTTTGTAGTGGCTCAGTCCAC 943
QY 989 AGTTTATCTCATGATTTTATTCGACTTTTCTATGTTTGAACAAATTTGAAGATTAGG 1048
Db 944 AATATATCTCTTGATATATATGCTCGATTTACACATGTTTGTATAGCTTCAAAAATTC 1003
QY 1049 AATTCACACATTTTCAGAGTGAATTTAAATTTTATGATGAAACATACATAGTCTTG 1108
Db 1004 AATTTCTCGGCTTTCAAGAGGAGATTTAGAACGCTATTTAGTGAACCTTTACAGGTTG 1063
QY 1109 GGTGGAACGAGATGAGCAATATTTTCATGGATGTTTGAACATGTTGCTTTTACGCTTT 1168
Db 1064 GATGGAAGGAGAGAAATATATTTCTTAGATGCTTCACTTTGTGCAATGGCTTTTGAAT 1123
QY 1169 ATTAAGGATCAATGGGTATGAAGTTTCCAGATTTCCAGATTTGGCTGAAATTTACTAA 1222
Db 1124 GTTACGGTTGAAGGATATGATGTTTCTTCAGACCAGTTTCACTCAATTTTCAAGATAT 1183
QY 1223 -----TGAATTAGCTTTGAAGAGCATATATGACGCTCTTGAACATATCA 1267

Db 1184 CTTTCCCAATTCCTTGGAGGATATTTAAAGACATTCGGTGCTCGCTGGAGTTATATAA 1243
QY 1268 TCGCTCACATATATATACCAAGAGGATTTACTTCT---GGAAACAATCTTGAAGTC 1324
Db 1244 GGCCTCTCAGATATACACGCCCGGATGAATCTGTTCTGGAAATATATAAATCTTTGGAC 1303
QY 1325 AGCTGATTTCTTCAAGAGATATATCCACTGATTCAAACAGGCTTCTAAA----- 1376
Db 1304 TAGTCGTTTCTTGAAGCATGATTAATCTAGTGATTCAGTTTGGTCTGATAGACCGATAG 1363
QY 1377 -TTAATTCACAAGAGGTGGAAATCTCTTAAGTTTCCCTATCAATACCGGTTTGAAGC 1435
Db 1364 TGTGTTTAAACAAGAGGCTTAAATCTCTTCTGAGTTCCCTATAATGCAACTCTAGAAGC 1423
QY 1436 CATATACTAGACAAATATATACAGCTTTACAAATGTAGACAATACAAGAAATCTGAAAC 1495
Db 1424 CCTAATAGTAGAGGGCAATGGAAGTTACAGTGGAGACATGTGAGGATTTCAAAATC 1483
QY 1496 TACATATCACTCATCAAAATATPAGTAACACTGATTACCTTAAGTTGGCTGTGTAAGATTT 1555
Db 1484 GCCATATCGCTTAAATTTTGGCCATCAAGATTTTCTGNACTTGTCTGATAGAGATTT 1543
QY 1556 CTACACCTGCCAATCTATTTATCTGTAAGAATTTAAAGTCTTGAAGGTGGTGGTGA 1615
Db 1544 CAATACCTGCAACGCATTTCTATTAAGAACTGGAAGAGCTTCAAGAGATGGTGGTGA 1603
QY 1616 GAATAAGTTGGACCAAGCTCAAGTTTCTAGGCAAAAGACCCCTACTGTTATTTCTCTGT 1675
Db 1604 AACAATTTGACGAGTTGAATTTTTCAGCTGCACTGAGGTACTGCTATTTTGTGTC 1663
QY 1676 TGTGCAACACTTTGCTCCCGAATATACAGATCGCGGTATTTCAAGGGCCAAAAATGG 1735
Db 1664 GCGAGCGACCTTACTGATCTGAACTTCATGATCTCGCATAGCTGCGGACAGGAATGG 1723
QY 1736 CATATTACTACAGTAGTTGATGACTTTTGTGATATCGTGGGTACAAATCGATGAATTCAC 1795
Db 1724 TGTGCTCACGACCGTGGTGTGATGTTTCTATGATGTTGGAGGATCTGAAGAGGAATGGA 1783
QY 1796 CAACCTGATTCAGATGTTGAAATGAAGTATGATGTCACAGAGGATTTGTTTCTAGA 1855
Db 1784 TAACTTATAGATTTGGTGAAGAGTGGGATCTGATGAGGAAAGTGGGTACTGTTCCAA 1843
QY 1856 GCATGTTTGGATTTTATTTTAGCATTTAAAGATGCAATCTGTTGGATTGGAGATGAAGC 1915
Db 1844 GGAAGTTGAGATTTGATTTCTTGCACTGCACAGCACAGTTTGTGAATAGGAAGAGAGC 1903
QY 1916 TTTTAAATGGCAAGCGCGATGTAACCTAGCCATGTTATTCAAACTTGGTTGGAACATAAT 1975
Db 1904 TTTAGTATGCAAGGACGACGTGTTATGAGGAATGTTATCGATGTTGGTGGCTCTGCT 1963
QY 1976 GAATAGTATGTTGAGAAGCTATATGCAAGAGATGCTTATGTCGCAATTAATGA 2035
Db 1964 GAAGGTGATGAGAAGAGAGCTGAATGTGCAAAATAGGTAGTGGCCATCAATGGGTGA 2023
QY 2036 ATATATGAAAACCTTACGTTGTCATTTGCAATTAGGCCCGATTTGTCAGCGCGCTATTTA 2095
Db 2024 ATATATGAAACAAGCCCATGATCATTCGCGTTGGACCTATATCTTCCAAATGCTCTT 2083
QY 2096 CTTTGTGGGCCCCAAATATCAGAGGAGATTTGTTGAAGCTCTGAAATATCATATCATTT 2155
Db 2084 CTTTGTGGAACCTTAACTCTCAGAGGAAATGATTTGGAAGCTGTGAATACCAAGATTTATA 2143
QY 2156 TAAGCTAATGAGCAGCGAGGTCCACTTCTAAGCATATCCATAGCTTCAAGAGGAAAT 2215
Db 2144 TAAGCTGATGAGCACTGCTGTGCTGCCCTTAAGATGATATTGCAATCTACGATAGAGATG 2203
QY 2216 TAAGGAAGGCAAAATTAACCGGTAGCATTTGCAATTTGAGTAAACGAGAAAGTGGGAAGT 2275
Db 2204 CAAAGAGGAAAGCTGAATATTTCTGTTGATGATTTGATGCGCGGTGATGTAC 2263
QY 2276 GGAAGAAGAGGTTGTGAGAGATGATGATGATTAATAAACAAGAGGAAAGATTAAT 2335

Db 2264 CAAAGAGGAGGCCATTTGAAGCAATTAAGGGGATTTTGAGAGGCGATTAAGAGCTGCT 2323
QY 2336 GAAATTAATTTTGAAGAAAATGGTAGCATTTCTTAGAGCTTTCTAAAGATGCAATTTTG 2395
Db 2324 GGGGTTAGTTTTCAGAGAGAACACTACAA---TTCCAAGAGCTTGTAGGATTTGTTCTG 2380
QY 2396 GAAACATGTCTACGCTGTTGAATTTTTCAGCAACGATGACGGTTTACTGGAAACAC 2455
Db 2381 GAAATTTGATGTCCTCAATGTCATTTTACATGGAAGATGATGGGTACATTTCAATAG 2440
QY 2456 GATTTCTGATCTGTGAAGGACATCATTTACAACCCGTTGGTGGCTTG 2502
Db 2441 GTTGTATGAACACTGTAAGGCCATCTTTGAACAACCCATGATCTGG 2487

RESULT 7
US-10-041-018-377
; Sequence 377, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 377
; LENGTH: 2506
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-041-018-377

Query Match 26.2%; Score 732.4; DB 16; Length 2506;
Best Local Similarity 59.5%; Pred. No. 9.2e-162;
Matches 1356; Conservative 0; Mismatches 876; Indels 48; Gaps 5;

QY 270 GGACAACTAATCCACTTAATCTGATCATTTGATACAAAGAACGGATCCAAACACAG 329
Db 143 GTACAGACAAGAGCTAACAATGTGAGCTTTGAGCAACAAGAGAGAGATTAGGAAGATG 202
QY 330 TTTAAATATGAGAAATTTCTTTTCTTATATGACACAGCATGGGTAGCCATGCTCCCT 389
Db 203 TTGAGAAAGTGGAGCTTCTGTTTCGCTCTACGATCTAGTTGGGTAGCAATGGTTCCA 262
QY 390 TCTCAAACTCACCACCAATCGCTTGTTCCTGAGTGTCTCAATTTGTTTAATTAAT 449
Db 263 TCACCGAGCTCCCAAAATGCTTCCACTTTCCACAGTGTGTGAATGGTTATTGGATAAT 322
QY 450 CAGCTTAATGATGTTTCATGGGTCTTGTGTTAATCACACTCATATCATATCAATCAAT 509
Db 323 CAACATGAAGATGATCTTTGGGACCTTGATAACC-----ATGACCATCATCTCTT 373
QY 510 CTTAAGATTTCTATCTTCAACATGATGATGATTTGCAATTAAGAGATGGAATGTT 569
Db 374 AAGAAGGATGTTTATCATCTACACTGGCTAGTATCTCGCTTAAAGAGTGGGAAT 433
QY 570 GGGGAAGATCAATAAATAAGGCTTAAGTTTATTTAGTCAATCTTGTCTAGCTACT 629
Db 434 GGTGAAGAATAAACAAGGCTCCAGTTTATTGAGCTGAATTTCTGATAGTCACT 493
QY 630 GAAAAAGTCAACCATCTCCCATTTGTTTGCATCATATTTCTGTTTGTGTTGAGTAT 689
Db 494 GATGAACCATACAGAAACCAAGGTTTGAATTAATTTCTGGGATGATTAATAT 553
QY 690 GCGAAATCTTGACATTAACCTCTCTTCAAAACAAACAGATTTTACTTTGATGCTACAT 749
Db 554 GCTAGAGATTTGAATCTGACGATTCATTCGCTCAGAGTGGGTGGATGACATGATACGA 613
QY 750 AAGAGGGAATTTGGACAAAAG-----ATGCCATTCAAATGAGATGGATGATAC 800

[illegible]

Db	1694	CTGGAACACCTCATACACTTGGTTCGAAAGTCGGGATTTGAAACGGTGTCTCTGAGTACAGC	1751
Qy	1851	TCGAGCATGTTCCGAAATTTTATTTTATAGCAATATAAAGATGCCAATCTGTGGATTTGAGAT	1910
Db	1754	TCGAAACATGTTGAGATCATATTTCTCAGTTCTTAAGGGACACCAATTTCTGAAACAGGAGAC	1813
Qy	1911	GAAGCTTTTAAATGGCAAGCCGCGAGTGAATAGCCATGTTATCAAACTTTGGTTGGAA	1970
Db	1814	AAAGCATTCACCTATCAAGGACGCAATGTGCACACCAACATTTGTGAAAATTTTGGTTGAT	1873
Qy	1971	CTAATGAATAGTATGTTTGAGAGAAGCTATATGGAACAGAGATGCTTATGTGCCAACATTA	2030
Db	1874	CTGCTCAAGTCTATGTTTGAGAGAAGCCGAGTGGTCCAGTGACAAGTCAACACCAAGCTTG	1933
Qy	2031	AATGAATATATGAAAACGCTTACGTCTCATTTTGCATTTAGGCCCGATGTCGAAGCCGGCT	2090
Db	1934	GAGGATTACATGGAATAATGCGTACATATCATTTTGCATTAGGACCAATTTTCTCCGAGCT	1993
Qy	2091	ATTTTACTTTTGTGGGGCCCAAAATATCAGAGAGAGATTTGTGAAAGCTCTGAATATCATAA	2150
Db	1994	ACCTATCTGATCGAACCTCCACTTCCAGAGAGACAGTCGATAGCCACCAATATAATCAG	2053
Qy	2151	CTATTTAAGCTTAATGACGACGAGGTCGACTTCTAAACGATATCCATAGCTTCAAGAGG	2210
Db	2054	CTCTACAAGCTCGTGAGCACTATCGGTTCGTTCTTAATGACATACAAGGTTTTTAAGAGA	2113
Qy	2211	GAATTTAAGGAAGCAAAATTAACCGCGTAGCATTTGACTTTCAGTACCGGAGAAAGTGGG	2270
Db	2114	GAAGCGCGGAAGGGAAGCTGAATCCGGTTTTTCATTCACATGAACACAGGAGAGACAAT	2173
Qy	2271	AAAGTGAAGAAGAGGTTGTGGAGGAGATGATGATGATTAATAAAACAAGAGGAAAGAA	2330
Db	2174	CGCAGCAAGAAGTGATCATAGAAATCGATGAAAGGTTTTAGCAGAGAGAAAGAGGCAAGAA	2233
Qy	2331	TTAATGAATAATTAATTTTGAAGAAAATGTTGATGATTTGTTCTTAGAGCTTTGTAAGATGCA	2390
Db	2234	TTGCAATAGCTAGTTTTTGGAGGAGAAAGGAAGTGTGGTTCCAGGGGAATGCCAAAGAGAGCG	2293
Qy	2391	TTTTTGGAACATGTCTCAGCTGTGCAATTTTTTTTTTACGCAACGATGACGGGTTTACTTGG	2450
Db	2294	TTCTTTGAAAATGAGCAAAAGTGTGGAACCTTATTTTACGGAAGGAGCATGGATTCACATCA	2353
Qy	2451	AACAGATCTTTGATCTGTGGAAGGACATCAATTTTACACCGGTTTGGTGCTTTGTGAATGAA	2510
Db	2354	AATGATCTTGATCAGTCTCTGTTTAAATCACTGATCTACAGAGCTGTTAGCTTACAGAAAGAA	2413

8 JUL 84

RESULT 8
US-10-425-114-3441
; Sequence 3441, Application US/10425114
; Publication No. US20040034888A1

: GENERAL INFORMATION:

APPLICANT: Liu Jingdong

APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jac

APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53313)B
TITLE OF INVENTION: Plants and Uses thereof
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28

Query Match 20.1%; Score 561.2; DB 16; Length 2554;

Best Local Similarity 56.1%; Pred. No. 2.1e-121; Matches 1210; Conservative 0; Mismatches 893; Indels 53; Gaps 6;			
QY	324	AAACAGTTTAAATCTAGAAATTTCTGTTCTTATATGACACAGATGGGTAGCCATG	383
Db	160	AATAGCTCCGGAAGCCCGAGTTGGCACCTTCTTATACACACAGCGTGGTTCATG	219
QY	384	GTCCCTTCTCCAAATCACCCAAATCGCCTGTGTTCCCTGAGTGTCTCAATTTGTTAAT	443
Db	220	GTGCCAGTGGGGCTCTCATCAGACTCCCGCTTCCCACTGTGTTGAGTGGATATG	279
QY	444	AATATACAGCTTAATGATGGTTCATGGGGTCT-TGTTAATCACACTCATATATCA	502
Db	280	CAGAAATCAGCAGGATCATGATCTTGGGGTGTGAGCGAATCTGACTCAT	328
QY	503	CCCGTTGCTTAAAGATTCTCTATCTTCAACATTAGCATGTATGTTGCAATTAAGAATG	562
Db	329	--CGTACAGAGGATGTTCTCTATCCACGTTGGCATGTGTTCTTGGTTGAAGAGATG	386
QY	563	GAATGTTGGGAAGATCAAAATAAATAAAGGCTAAAGTTTATTTAGTCAAAATCTTGTTC	622
Db	387	GAATGTTGGCAGAGACATTTTGGAGAGACTGCAATTCATCGGAGGAATTTCTCTGT	446
QY	623	AGCTACTGAATAAAGTCAACCATCTCCCATTTGCAATTTTTCATCATTTCTCTGTTGCT	682
Db	447	TGCTATGGACCAAGTTCACCTTCTCTATAGGTTTCAACATCACCTTTCTCTGTTGCT	506
QY	683	TGAGTATCGGAAACTTGGACATAAACCTCTTTCATAAACAACAGATTTTAGTTGAT	742
Db	507	TAACTCGGCATTGATATGGGTTAGAAATTTCTGTAAAGACAAACTGATGTCTGTTGGCAT	566
QY	743	GCTACATAAGAGGAATTTGGAGCAAAAAGATGCCA-----TTCAAATGAGATGA	793
Db	567	TCCTTACCGCGGAGATGGAATTCAAAAGGCTGCTGGATAGTCTCTTTTGGAGAAA	626
QY	794	TGATACTTGGGTATATCTCTGAGAGACTCGTAAATTTATGATTTGGAATATGGTGAA	853
Db	627	AGCATATATGGCTTTTATCCAGAAGGATTCGGAATATGCTGGACTCGGATCAAGTTAT	686
QY	854	GAATATCAGATGAATAAGTCTGTTTTCATCTCACCATCAGCAACAGCTGCTGCTTT	913
Db	687	GAAGTTTCAGAGGAAGAAATGATCATTTGTCAGCACTCTTCCCAACTGCTGTTGCAT	746
QY	914	CATTAAATCATCAAATCTGTTGTTCTTAATTTAATTTCACTTTTGGACAAGTTTGG	973
Db	747	AATCCACAAATACACGACCAGCCCTTCATACCTTAATTTGTTGTGAGTGAATTTGG	806
QY	974	TAATGCAATCCCAACAGTTTATCTCATGATTTTATTCAGACTTTCTATGTTGACAC	1033
Db	807	CAGTGCAGTACAGCAATGTATCTTCAATGGTACATTTGTGAGCTTTCAGTGAATTTG	866
QY	1034	AATTGAAAGATTAGAAATTTCCACACCATTCAGAGTGGAAATTAATAATGTTTAGATG	1093
Db	867	GCTTGAATAAATGGAATTTCTCAGCGTTTGTGAGTGAATAAGAAAGCAATCTGGA	926
QY	1094	AACATACAGATGTTGGTGGAAACAGATAGCAATAATTTCAATGATGTTGTAACATGTC	1153
Db	927	GGCATACAATTTGCTGTTTACAGAAAGATGAGAAATCATGATGACATAGCAACATGTC	986
QY	1154	TTTACCTTTCCGTTTATAGGATCAATGGGTATGAAAGTTTCCACAGATCCATTTGGCTGA	1213
Db	987	AATGGCAATTTCCGCTTTTGGAGTGAATGGTTTACATGTTTCTCTCAGATGATGTTCTCA	1046
QY	1214	AATTACTTAATG-----AATTAGCTTTGAAAGACGAATATG	1249
Db	1047	CGTTGCTGGAGCTTCCACTTTCCATGATTCATCAAGAGGATTTTAAATGATACAAAATC	1106
QY	1250	AGCTCTTTGAAACATATCATCGTGCACATATATATATACCAAGAGATTTTATCTTCTGAAA	1309
Db	1107	CCTACTGGAATTTGACAGACCTCAAGTCAACCTTATCAGAAACGATCTGATCTTAGA	1166
QY	1310	ACAAATCTTGAGTCACTGATTTCTTCAAAGAGATAATATCCACTGATTTCAAACAGGCT	1369
1167	TCGCATAGTTCCTGGTCTGGCAACTTATTGAAGGATTAAGATGTGCTGTAGTAGGGTGA	1226	
1370	TTCTAAATTAATTCACAAAGAGGTGGAAATCTCTTAAGTTCCCTATCAATACCGGTTT	1429	
1227	AAAAGACTCGAATTTTGGAGAGATGGAGTATGCTGTAAATTTCCCTTGTATCCACACT	1286	
1430	AGAACGCATTAACACTAGACGNAATATACAGCTTTACAAATGTAGACAATAAAGAAATCT	1489	
1287	GGAGCGTCTAGAACACAAAGAGAAACATCGAACATTTTGATGCTTGGGGTCTCTCATGCT	1346	
1490	GAATACTACATATCACTCATCAAAATATTAGTAAACACTGATTACCTAAGCTTGGCTGTGA	1549	
1347	---AACACAAATCTCTCATCTTTTCGTATCAATCAAGAAATTCCTAGCTTGGCAGTGA	1403	
1550	AGATTTCTACCTCCCAATCTATTTATTCGTGAAGAATTAAGAAGCTCTTTGAAGGTGGT	1609	
1404	AGATTTCAATTTCTCTCAACGTTGTTTACCGGATCGAACTTCGGCATCTTTGATGTTGGT	1463	
1610	GGTAGAATTAAGTTGGACACAGCTCAAGTTTGTCTAGGCAAAAGACCGCTACTGTATTT	1669	
1464	GAAGAGAAACAGCTGGACCACTACAAATTTGCTCGGCAAAACCTGACATATTGCTATCT	1523	
1670	CTCTGTTCTGCAACACTTTCTGCTCCGAAATTAATCAGATGCGGTATTTTCATGCGCCAA	1729	
1524	GTCTGCTGCTGCTACCGTATTTTCTTGAATTTGCTGACCTCGCATTCGATTCATGGCCAA	1583	
1730	AAATGGCAATTAACACTACAGTATGATGACTTTTTTGTATATCGGTGGTACAAATCGATGA	1789	
1584	AAATGGTGTCTCAACTGCTGGTTGATGACTTCTTCGATGTGTGTGATCAAAAGAAGA	1643	
1790	ATTGACCAACCTGATTCATGTTGTTGAAAATGGAATGTAGATGTCGACAAGGATGTTG	1849	
1644	ATTGAAAACCTGTATAGCACTAGTTGAGAAATGCGCATGCGCACCATGCAATGTGATCTA	1703	
1850	TTCAAGACATGTTCCGATTTTATTTTAGCAATTAAGAATGCAATCTGTTGATTTGAGA	1909	
1704	TTCGAACAGTGAATATGATTTTCTGCTATTTTATACACAGTGAACCATCTTTGGAGC	1763	
1910	TGAAGCTTTTAAATGGCAAGCGCGATGTAACCTGTAACCTGTTATTTCAAACCTGTTGGA	1969	
1764	AATGGCTTTCAGACAACAAGCCGTTGATTTTACAAACCACTAGTAGAATAATGGCTGA	1823	
1970	ACTAATGAATAGTATGTTGAGAGAGCTATATGACAGAGATGCTTATGTGCCAACAT	2029	
1824	TTTGTTAAGATCTATGATGTTGCGAGGAGAAATGCGAGAGATGCCAATATGTACCAACAGT	1883	
2030	AAATGAATATATGAAAACGCTTACGTTGCTCAATTTGCAATTAGCCCGCTGTTCAAGCCGC	2089	
1884	TGAAGATACATGACAAATGCTGTTGCTCATTTGCACTGGGCCCAATTTGCTCCACGC	1943	
2090	TATTTACTTTGTTGGGGCCCAATTTATCAGAGGAGATTTGTTGAAAAGCTCTGTAATATCA	2149	
1944	ATTGTATTTTGTAGGGCAAGAGCTATTATAGACATGCTGTCAAAGATGAAGATACGATA	2003	
2150	TCATTTAAAGCTAATGACGCGAGGCTGACCTTCTTAAAGATATCCATAGCTTCAAGAG	2209	
2004	ATTTATTTAGCTAGTAGACCTTTCGGGAGGCTCTCTCAATGACTACCAAAAGTTTAGAG	2063	
2210	GGATTTTAAAGAGGCAAAATTAACGCGGTAGCAATTCATTTGATTAACGAGAAAGTGG	2269	
2064	GGAGGCAACCGGGGAGCTGAAATGTTTCTTACTTGTGCTCCACAGTGTGTTGCTT	2123	
2270	GAAAGTGGAAAGAGGTTGTGGAGGAGATGATGATGATGATTAATAAACAAGAGGAAAGA	2329	
2124	TATGTCCTAGAGACCGCTAAAAGGCAATGACAGAGTCCATAGACGCTGCTAGGAGAGA	2183	
2330	ATTAATGAATTAATTTTGAAGAAATGCTAGCATTTGTTCTTAGAGCTTTGTAAGATGC	2389	
2184	CTTGTGAAGATTTGGTTCT---CAGGAAGAAAGTGTCTTTCTTAGGCAATGCAAGAGCT	2240	
2390	ATTTTGGAAACATGTCACGTTGTGAATTTTATTTTGTAGCAACGATGACCGGTTTA	2445	
2241	CTTCTGGAAGATGTGAAGTACTTCACTGTTTTTACTCTCAGATGATGGAATTTA	2296	

RESULT 9
US-10-425-114-17004
; Sequence 17004, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 17004
; LENGTH: 2086
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3067-045-E3_FLI
US-10-425-114-17004

Query Match 15.5%; Score 433.2; DB 16; Length 2086;
Best Local Similarity 55.8%; Pred. No. 2.7e-91;
Matches 947; Conservative 0; Mismatches 713; Indels 36; Gaps 5;

QY 778 ATTCAATGAGTGGATGATCTGCGTATATCTCTGAAGACTCGGTAAATTTATATG 837
DB 119 AATCTAGCGCAATAGAGCATATTTGCCATGTTCTGAGAGGTTCGAAACCTCTGG 178
QY 838 AITGGAATATGTTGAAGAAATATFAGATGAAAAATGGTTCTGTGTTTCAACTCACCATCAG 897
DB 179 ACTGGAATGATGTTATGAAGTTCCAAAGCGAAGATGGATCCTGTTTAACTCTCCTCTG 238
QY 898 CAAAGCTGCTGTTTCAATTAATCATCAAAATCCTGGTGTCTTAATTTAAATTCAC 957
DB 239 CAACTGCTCGCTTGGTGGCCAACTATGACGACAAAGCGCTACAGTATCTAAATTTGC 298
QY 958 TTTTGGACAGTTGTTGTAATGAGTCCCAACAGTTTATCTCATGATTTATTTATCCGAC 1017
DB 299 TTGTCACAAATTTGGAGTGCAGTACCAACAGTTTCCCAAAATTTCACTATCAGC 358
QY 1018 TTTCTATGTTGACAAATTTGAAGATTTAGGAATTTTACACCAATTTTCAAGTGGAAATTA 1077
DB 359 TTTCAATGTTGGACACGCTCGAAAGTTTGGAAATATCAAGGCAATTTTCTGTGGAGAAA 418
QY 1078 AAAATGTTTATGATGAAACHATGACAGATGTTGGTGGNACAGATGACCAATTTATCATGG 1137
DB 419 AGGCTGTCTCGACATGATATACAGTTTGTGGTTACAGAAAGAGGAGGAATATGTTGG 478
QY 1138 ATGTTGTAACATGCTGTTTACGCTTTCGGTTATTAAGCATCAATGGGTATGAAGTTTCCC 1197
DB 479 ATCGAAGAACATGTCGATGCGATTTTCCGAATTTTCAAAATGATGGCTTTGATGTTTCCA 538
QY 1198 CAGATCCATTTGGCTGAAATTTACTAATG-----AATAGCTT 1233
DB 539 CAGATTTGTTGTTCTCATATTGTTGAAGCTTCCAAATTTTCCAACTCACTACAAAGGATAT 598
QY 1234 TGAAGACGATATGAGCTCTTGAACATATCATGCTC---ACATATATTATACCAAG 1290
DB 599 TATGTGATCAAAAGACTCTATTGGAATGATGACAAAGCTCTTAAAGTCAATATTATCAGAAA 658
QY 1291 AGGATTTATCTTCTGGAAGAACAAATCTTGAAGTCAAGTCAAGTCTTCTCAAAGAGATAATAT 1350
DB 659 GAGACTTCACTTGAAGAACATAGATGTTGACAGGTAGCTTATTGAGGAGAAAGCTAT 718
QY 1351 CCAGTGTATCAACAGGCTTTCTAAATTAATTTCAAGAGGTGGAAGAAATGCTCTTAAGT 1410

DB 719 GCTCTGATGGA---CGCAAAAGAGTACCGATATCTTGAAGAGTGAAGTATGCCCTTAAGT 775
QY 1411 TCCCTATCAATACCGGTTTAGAAACGATATAACACACTAGACGAAATATACAGCTTTTCAATG 1470
DB 776 TTCCCATTTATGCCCTTAGTAGATCTCTCAACAACAGAGAGATGATGGCAATTTGGATG 835
QY 1471 TAGACAATACAGAATTTCTGAAGAACTACATATCATCTCAATCAATATTAGTAACATGATT 1530
DB 836 CTAGGGGTTCTCAGACGCTTAAAGACAAAATATCTCCATGTCTCATGTAACTCA---AGACA 892
QY 1531 ACCTAAGGTTGGCTGTTGAAGATTTTACACCTGCTCCATCTATTATTTCTGTAAGAAATTA 1590
DB 893 TCCTTGGCTGGCTGTTGAAGATTTTCTGTTTCTCAATCTATTATCTACCGAGAGCACTAC 952
QY 1591 AAGGCTTTGAAAGTGGGTGAGAGATAAGTTGGACCCAGCTCAAGTTTGTAGGCAAA 1650
DB 953 AGAATCATATTAGTTGGGAGAAAGAGATAGGATGGACCTACAAATTTGTGCGGCAAA 1012
QY 1651 AGACCGCTTACTGTTTATTTCTCTGTTGCTGCAACTTTTCTGCTCCCGAATTTATCAGATG 1710
DB 1013 GGCTGGCATATTGCTATCTCGCTGCTGCTACCACCATATCCCTCATGAATTTGCTGTGATG 1072
QY 1711 CGCGTATTTATGCGGCCAAATATGGCATATTACTACAGTAGTTGATGACTTTTGTGATA 1770
DB 1073 CTCGGTTGCTATGTCGCAAAAGCATCATGCTCAGGTTGATGTTGATGACTTTCTTCGATG 1132
QY 1771 TCGGTGTACAAATCGATGAATTCACCACTCATCTCAATGTGTGTAATAATGGAATGTAG 1830
DB 1133 TTGTTGATCAAAAGNAGAACAGAAATCTCATCGAATTAGTTGAGAACTGGATGAGC 1192
QY 1831 ATGTCGAAGAATTTGTTCTCAGAGCATGTTCTGGATTTTATTTTATTTAGCATTAAGAATG 1890
DB 1193 ACCACAAAGTTGAGTTCTCTGCGGAGAAAGTAAATAGTTTCTATGCTGTCTATAATA 1252
QY 1891 CAACTCTGTTGGATGAGATGAAGCTTTTAAATGCAAGCGCGCATGTAACTAGCCATG 1950
DB 1253 CAGTGAACAGCTTTGGATCTATGGCTTTCTGAGTACAGAGCCGATGTCACAAACACC 1312
QY 1951 TTATTTCAAACTTTGGTTGGAATTAATGAATAGTATGTTGAGAGAACTATATGACCAAG 2010
DB 1313 TCGCTGAACATGGCTTAAAGTAATGCTGTGATGCTGACGAGGAGGAGAGCTGCGCAAGGA 1372
QY 2011 ATGCTTATGTCGCAACATTTAATGAATATATGAAACCGCTTACGTTGCTATTTGCAATTAG 2070
DB 1373 GGCATTTGTTACCAACAGTTGAGGAATATATGCAAAATGCAAGTTTCTCTGCGCACTGG 1432
QY 2071 GCCCGATTTGTCAGCGGCTATTTACTTTGTTGGGCGCCAAATTTATCAGAGGAGATTTG 2130
DB 1433 CGGTCAATTATATCTTCAGCGCGCATTTTCTTGGCGAAACGCTCTCAGATTAATGTTCA 1492
QY 2131 AAAGCTCTGAATATCATATCTATTTAAGCTAATGAGCAGCGAGGTTGCACTTTCTTAAACG 2190
DB 1493 AAGATCATGAATACAGCAATTTTCTGAGCTCATGTTCACTTCAGTCTGCTCTTAAATG 1552
QY 2191 ATATCCATGCTTCAAGAGGAAATTTAAGAAAGCAAAATTTAAACGCGGTAGCATTTGCAAT 2250
DB 1553 ACATTCGAAGCTTGAAGGAGGAGTTGCAAGCTGTTAACTGTAACAGCGTTTCACTGCTTG 1612
QY 2251 TGAGTAAAGGAGAAAGTTGGGAAAGTGGAAAGAGGTTGTGGAGGAGATGATGATGATGA 2310
DB 1613 CTCTACAGTGGTGGTTCCATGTCATAGAGCGCTAAGAGGAAACACATGATCTT 1672
QY 2311 TTAAGAACAGAGAGAAAGATTAATGAATTAATTTTGAAGAAATATGTTAGCATTTGTTTC 2370
DB 1673 TAGAGTCGTATAGGAGAGGTTTAGTACGTTGGTTGCT---AGACAGGACAGTGTGTTTC 1729
QY 2371 CTAGAGCTTGAAGATGCAATTTTGGACATGTTGACGTTGTTGAATTTTATTTTACGCAA 2430
DB 1730 CTAGGTCAATCAAGGAGCTGTTCTGGAAGCTTTTGAAGCAATTTCACTGTTTCTACTCC 1789
QY 2431 ACGATGACGGTTTAC 2446
DB 1790 AGATTGATGATTTAC 1805

RESULT 10
US-10-425-114-5988
; Sequence 5988, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 5988
; LENGTH: 2029
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700550363_FLI
US-10-425-114-5988

Query Match	14.1%;	Score 392.4;	DB 16;	Length 2029;
Best Local Similarity	53.9%;	Pred. No. 1.1e-81;		
Matches	999;	Conservative 0;	Mismatches 741;	Indels 112; Gaps 5;
QY	700	TGGACATAAACCTCCTTTCAAAACAAACACAGATTTTAGTTGATGCTACATAGAGGGGAT	759	
Db	4	TGGGTTTAGAATTTCTGTAAACAAACTGAAGTCTGTGGCATTTCTTACCGCCCGGAGA	63	
QY	760	TGGAGCAAAAAG-----ATGCCAATCAAATGAGATGGATGCATCTTGGCGGTATA	810	
Db	64	TGGAATTTGAAAAGGCTGCCTGGTAGTTCTTTTGGGAAGAAAGCATATATGGCTTTTA	123	
QY	811	TCCTGGAAGACTCGGTAATTTATATGATTTGGATATGTTGGTGAAGAAATATCAGATGAAAA	870	
Db	124	TCCCAGAAGGATTCGGAAATATGCTGGAGCTGGGATCAAGTTATGAAGTTTCAGAGGAAGA	183	
QY	871	ATGGTTCTGTGTTTCCAACCTCACCATCAGCAACAGCTGCTGCTTTCATTAATCATCAAAATC	930	
Db	184	ATGGATCATTTGTTCCAGCACTCTTCCCACTGCTGTTTGCATTAATCCACAAATACACG	243	
QY	931	CTGGTTCTCTTAATATTATTAATTTCACTTTTGGACAAGTTTGGTAATCAGTCCCAACAG	990	
Db	244	ACCAAGCCCTTCAATACCTTAATTTGCTTGTCAAGTGAATTTGGCAGTGCAGTACCAGCA	303	
QY	991	TTTTATCCTCATGATTTATTTATCCGATTTTCTATGGTTGACACAAATGGAAGATTAGAA	1050	
Db	304	TGATCCTCTCAATGGTACATTTGTCAGCTTTCAATGGTGGACGCGCTTGAAGAAAATGGGAA	363	
QY	1051	TTTTCACACCACTTTCAGAGTGGAAATTAATAATGTTTTAGATGAAACATACAGATGCTCGG	1110	
Db	364	TTTTCTCAGCGCTTGTCAAGTGAATAGAAAGCATCCTTGGACATGGCATACAAATGCTGGT	423	
QY	1111	TGGAAACGAGATGAGCAAAATATTCATGAGATGTTGTAACATGTCGTTTAGCTTTTCGGTTAT	1170	
Db	424	TACAGAAGATGAGGAAATCATGATGAGCATAGCAATGTCGAATGGCAATTTGCGCTTT	483	
QY	1171	TAGGATCAATGGGTATGAAGTTTCCCGAGATCCATTTGGCTGAAATTAATGAATAGTAG	1230	
Db	484	TGAGGATGAATGGTTTCAATGTTTCCCTCAGATGAGCTGTCTCAGTTGCTGGAGCTTCCA	543	
QY	1231	CTTTGGAAGACGATATGCACTCTTGAAA-----	1260	
Db	544	CTTTCCATGATTCACACAGGATTTTAATATGATACAAAATCCCTACTCGGAATGTGACA	603	
QY	1261	-----CATATCATCGCTCACATATATTATATCCA	1288	

604	DB	AGACCTCAAAGTCACCTTATACAGAAAACGATCTGATCTTTAGATCGCATAGGTTCTCTGGT	663
1289	QY	AGAGGATTTATCTTCTCGGAAAACAAATCTTTGAAGTCAGCTGATTTCTCTCAAAGAGATAAT	1348
664	DB	CTGGCAACTTATTCGAAGGATAAGATGCTGTAGTAGGGTGCAAAAAGACTCGATTTTGTG	723
1349	QY	ATCCACTGATTCAAACAGCGCTTCTTAAATTAATTCACA-----	1386
724	DB	GAGAGGTGCTGCAAAAAAAATTTAAATCTCAATTTCACTGGAGGTTTCAGTTTATGTC	783
1387	QY	-----AAGAGGTGGAAAATGCTTTAAAGTTCCTATCATATCGGTTTAGAA	1433
784	DB	TGTGTTTTGTGATTTTCAGATGGAGTAGCTGTAAATTTTCCCTTGTAATCCACACTGGAG	843
1434	QY	CGCATAAACACTAGACGAAATATACAGCTTTACAATCTAGACAATACAAGAAATTCGAAA	1493
844	DB	OGTCTAGAACACAAGAGAAACATCGAACATTTTCATGCTTGGGGTTCCTGTATGCT---A	900
1494	QY	ACTACATATCACTCATCAAAATATATAGTAAACACTGATTAACATAAGTTGGCTGTGAAGAT	1553
901	DB	ACAAACAAATTCCTCATCTTTTCGTATCAATCAAGAATTCCTAGCTTTGGCAGTCGAAGAT	960
1554	QY	TTCTACACCTGCCAATCTATTATTCGTPGAGAATTTAAAGGTCCTGAAAGTTCGGTGGA	1613
961	DB	TTTCAGTTTCTCTAACTGTGTTACCGGGATGAACCTCGGCATCTTGATAGTTCCGGTGAAG	1020
1614	QY	GAGAAATAGTTGGACCAAGCTCAAGTTTGTAGGCAAAAGACCGCTACTGTTATTTCTCT	1673
1021	DB	GAGAAACAGCTGGACCAAGCTACAATTTGCTCGCAGAAACTGACATATTCATCTGTCT	1080
1674	QY	GTTGCTGGAACACTTTCGCTCCCGAATATACAGATCGCGTATTTTCATGCGGCCAAAAAT	1733
1081	DB	GCTGCTGCTACCGTATTTTCTCTGAATTTGCTCGCGTCCGATTCGATAGTTCCGCTCT	1140
1734	QY	GGCATATTAACACTAGTAGTTGATGACTTTTTTGTATTCGGTGGTACATCGATGAATG	1793
1141	DB	GGTGCTCTCAACTGTGGTGTGATGACTTCTTCGATGTTGGTGGATCAAAAGAAGAAATTA	1200
1794	QY	ACCAACCTGATTCAAATGTGTTGAAAATGGAATGTAGATGTCGACAAAGGATTTGTTTCA	1853
1201	DB	GAAAACCTGATAGCACTAGTTTGAATGCGATGCGCACCATGCACTGAGTTGATTTCTATTCG	1260
1854	QY	GAGCATGTTCCGATTTTATTTTATAGCAATTAAGAATGCAATCTGTGATTTGGATTTGGAGTAA	1913
1261	DB	GAAACAGTGAAAATAGTATTTTCTGCTATTTATACAAACAGTGAACCAATCTTGAGCAATG	1320
1914	QY	GCTTTTAAATGCGAAGCGCGCATGTAACTAGCCATGTTATTCAAACTTGGTTGGAACTA	1973
1321	DB	GTCTCTGCGACAAGGCGGTGATTTTACAAACCACTAGTAGAAATATGGCTGGATTTG	1380
1974	QY	ATGAATAGTATGTTGAGAAAGCTATATGCAAGAGATGCTTATGTGCGCAACATTAAT	2033
1381	DB	TTAAGATCTATGATGGTCGAGGAGAAATGGCAGAGATGCCAATATGTACCAACAGTTGAA	1440
2034	QY	GATATATGAAAACCGTTACGTGTCAATTTGCATATAGGCCCGATGTTCAAGCCGGCTATT	2093
1441	DB	GAATACATGACAAATGCTGTGTCTCATTTGCACTGGGCCCAATTCGTCTCCAGCATTTG	1500
2094	QY	TACTTTTGTGGGCCCCAAATTTATCAGAGGAGATGTTGTAAGACTCTGAATATCATATCTA	2153
1501	DB	TATTTTGTAGGGCAAGAGCTATTAGAGCATGCTGTCAAAAGATGAAGAGTACGATAAATTA	1560
2154	QY	TTTTAAGCTAATGAGCAGCGAGGTCCACTTCTTAAACGATATCCATAGCTTCAAGAGGAA	2213
1561	DB	TTTTAGGCTAGTAGCACTTTCGGGAGGCTCCTCAATGACTACCAAAAGTTTAGAGAGGHA	1620
2214	QY	TTTTAAGGAAGCAAAATTAACCGCGTAGCATTTTCATTAACGAGAAAGTGGGAAA	2273
1621	DB	GGCAACACAGGGGAAGCTGAATAGTGTCTCTACTTTGTCTCCACAGTGTGTGTTCTATG	1680
2274	QY	GTGGAAGAAGAGGTTGTGGAGGAGATGATGATGATTAATAAAACAAAGAGGAAGAAATTA	2333
1681	DB	TCCATGAGCCGCTAAAAAGGCAATGCGAGAAGTCCATAGACGTGTCTTAGAGAGACTTG	1740

QY 2334 ATGAAATTAATTTTGAAGAAAATGGTAGCAATGTTCTTAGAGCTTGTGAAAGATCAATTT 2393
Db 1741 CTAAGATTGGTCT---CAGGAAGAAAGTGTCTGTTCTTAGGCCAAGCAAGAGCTCTTC 1797
QY 2394 TGGAACTGTGTGACGTGTGAAATTTTACGCAAAAGATGACGGGTTTA 2445
Db 1798 TGGAAATGTGTAAAGATCTTACCTGTTTACTCTCAGAATGATGATTTTA 1849

RESULT 11
US-10-259-194A-107
; Sequence 107, Application US/10259194A
; Publication No. US20040010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Ghassenian, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Riche, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
; CURRENT APPLICATION NUMBER: US/10/259,194A
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,743
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 662
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 107
; LENGTH: 2223
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-259-194A-107

Query Match 13.6%; Score 379.4; DB 16; Length 2223;
Best Local Similarity 54.5%; Pred. No. 1.3e-78;
Matches 900; Conservative 0; Mismatches 676; Indels 75; Gaps 4;

QY 795 GGATACCTTGGCGTATATCTCTGAAGGACTCGGTAAATTTATATGATTTGGAATATGGTGAAG 854
Db 574 GAATATACGGCTTAUTCTCTCAGAGGATTTAGGCAATTTTCAGNACTGGATGGAATGATG 633
QY 855 AAATATCAGATGAAATGGTCTCTGTTTCACTCACTCACTCACTCACTCACTCACTCACTCACT 914
Db 634 AAATCCAAAGAAAGATGGCTCACTGTTTCACTCCCTTACACAACTGCACTGCAATTA 693
QY 915 ATTATCATCAAAATCTGGTGTCTTAATTTAAATTTCACTTTTGGACAGTTTGGT 974
Db 694 GTCCCAACTATGATGCCAAAGCTCTCCAGTACTTAGACATGTTCTTGACAAATTTGGA 753
QY 975 AATGAGTCCCCAACAGTTTATCTCATGATTTATTTATTTATTCGCACTTTCTATGGTTGACACA 1034
Db 754 AGTGAGTCCAGCGGCTTATCTGCAATTTTCAGTCTCAGCTCTACATGTTGGATG 813
QY 1035 ATTGAAGATAGGAATTTCAACCACTTCAGAGTGGAAATTTAAATGTTTATGATGAA 1094
Db 814 CTTGAAAGATGGGAATATCTAGGCATTTTGTGTGATGATGATGATGATGATGATGATGATG 873
QY 1095 ACATACAGATGTTGGTGGNAAGAGATGAGCAATTTATCATGATGTTGTAACATGCT 1154
Db 874 ACCTACAGTTGCTGGAAACAGAGGATGAGAAATTTGCTTGACATGCAATGTTGGG 933

QY 1155 TTAGCTTTCCGTTTATTAAAGATCAATGGGTATGAAGTTTCCCGAGATCCATTTGGCTGAA 1214
Db 934 ATGGCAATTTCCGTTTATGATGAATGGATGATGATGTTTCTTTCAGATGAGCTATCTCAT 993
QY 1215 ATTACTAATGAATTAGCTTTGAAGAGCAATATGACGCTTTGAAGCAATATCATCGTCA 1274
Db 994 TTTTCT-----GAACTTCAAGTTTCCCAATTTCA 1023
QY 1275 CATATATTATACCAAGAGATTTATCTTCTCGAAAAAATAATCTTGAAGTCAAGCTGATTT 1334
Db 1024 CTTCAAGGATATCTGAATGATACAGATCTTTATTAGATTTACATAAGGCTTCAAAAGTC 1083
QY 1335 CTCRAAGAGATAATATCCACTGATTTCAAAACAGGCTTTCTAAATTAATTTCAAAAGAGTG 1394
Db 1084 -----AGTATCCGCAAAAAGAGGTTT 1104
QY 1395 GAAATGCTCTTAAGTTCCCTATCAATACCGGTTTGAACGCAATAAAGCACTAGACGAAT 1454
Db 1105 GAATATGCTCTTGAATTTCCCTTCTATACCACTTTGGATCGTCTAGACCATAAAGAAAT 1164
QY 1455 ATACAGCTTTTACAATGTAGACAATAACAAGAAATTTCTGAAAACTTACATATCATCTCAAT 1514
Db 1165 ATCGAAGATTTTGACATTTACAAGCAGTCAGATGCTTAGAACAAGCT---ACTTGCATCT 1221
QY 1515 ATTAGTAACTGATTA CTTAAGTTGGTGTGTTGAAGATTTCTACCTGCCAATCTATT 1574
Db 1222 CATTCCAATGAAGAAATCATGGCTTGGGTGTGAGAGATTTTATAGTCTCTCAGTTTATT 1281
QY 1575 TATCGTGAAGATTAAGAGGCTTTGAAAGTGGGTGGTAGAATAAGTTGACACAGCTC 1634
Db 1282 TTCCAAAGAGAGCTCAGCAACTCAACAGCTGGGTGAAAGAGAGAGGTTGGATCAGCTG 1341
QY 1635 AAGTTTGTAGGCAAAAGACCGCTTACTGTTTATTTCTCTGTTTGTGCTGCAACACTTTCTG 1694
Db 1342 CAATTCGACGCGCAGAGTTGACATCTTCTTATTTCTCTGCTGCTGCTACCACTTTTCACT 1401
QY 1695 CCCGAATTAATCAGATGCGCGTATTTTCATGGGCAAAATATGCGATTAATTAATCAGTAGTT 1754
Db 1402 CCTGAACTGTGAGATGTTTCGCAATTTTGTGGGCCAAAAATGGCGTGTGCAACAGGTCGTC 1461
QY 1755 GATGACTTTTGTGATATCGGTGATCAATCGATGATTAATGACCACTGATTTCAATGTTT 1814
Db 1462 GACGACTTCTTCGAGCTTGGAGGATCAAAAGAAAGACTGGAATACTCTGCCATTTAGTT 1521
QY 1815 GAAAAATGGAATGTAGATGTCGACAAGGATTTGTTTCAGAGCATGTTCCGATTTTATTTT 1874
Db 1522 GAGAGTGGGCAAGAAATGACAAACTGAGTACTACTCTGAAACAGTAGAGATTGTGTTTC 1581
QY 1875 TTAGCAATTAAGAGATGCAATCTGTTGGATTTGAGATGAGAGCTTTTAAATGGCAAGCGCG 1934
Db 1582 TCTGCAATTTATACCTTCACTAACCCAGCTTGGATCAATGGCCCTCTGTAGTACAGGCCCT 1641
QY 1935 GATGTAACCTAGCCATCTTATTTCAAACTTGGTTGGAATTAATGAATAGTATGTTGAGAGAA 1994
Db 1642 GATGTCACCAACACCTTGTAGAAATATGGCAAGAAATGCTGAGGCTTAATGATGACAGAG 1701
QY 1995 GCTATATGACAGAGATGCTTTATGTGCCAACAATAAATGAATATATGAAAAACGTTTAC 2054
Db 1702 GTAGAGTGGAGCAGAGCCGGTATGTGCCAACAGCAGAGGAATATCATGAAAAATGCAATT 1761
QY 2055 GTGTCAATTTGATTTAGCCCGGATTTGCAAGCCGGCTATTACTTTTGGGGGCCCAATTA 2114
Db 1762 GTGACATTTTGCATGGGACCCCGTTGTCTCCAGCATTTGTATCTTCTTGGACCAAGATC 1821
QY 2115 TCAGAGGAGATTTGTAAGAGCTCTGAATATCATATCTATTATTAAGCTAATGAGCAGCGCAG 2174
Db 1822 CTTGATCTGTCAATAGAGCCAGAAATGACCGAGTGTGTTCCGGCTAATGAGCAATGT 1881
QY 2175 GGTTCGATTTCTAAACGATATCCATAGCTTTCAAGAGGAAATTTAAGGAAGGCAAAATTAAC 2234
Db 1882 GGCGGTCTCTGATGATGTCCTCAATCTCTACGAGAGAGGCGCAGCCAGGCAAGCTGAAC 1941
QY 2235 GCGGTAGCATTTGCATTTGATTAACGAGAGAAAGTGGGAAAGTGGAGAGAGGTTGTGGAG 2294

RESULT 13

US-10-041-018-375

; Sequence 375, Application US/10041018

; Publication No. US20040072323A1

; GENERAL INFORMATION:

; APPLICANT: Mateuda, Seiichi P.T.

; APPLICANT: Hart, Elizabeth A.

; TITLE OF INVENTION: diterpene-producing unicellular organism

; FILE REFERENCE: P02080US1/10025547

; CURRENT APPLICATION NUMBER: US/10/041.018

; CURRENT FILING DATE: 2002-01-07

; PRIOR APPLICATION NUMBER: US 60/259880

; PRIOR FILING DATE: 2001-01-05

; NUMBER OF SEQ ID NOS: 413

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 375

; LENGTH: 1559

; TYPE: DNA

; ORGANISM: Zea mays

US-10-041-018-375

Query Match 11.8%; Score 330; DB 16; Length 1559;

Best Local Similarity 55.7%; Pred. No. 4.9e-67;

Matches 655; Conservative 0; Mismatches 515; Indels 6; Gaps 1;

QY 1270 CGTCACATATATTACCAAGAGCATTTATCTCTCGAAAAACAAATCTTGAAGTCAGCTG 1329

DB 116 CGACAGTTAGTATCTCTGAAGTAGGCTATCTCTGGATAGCATAGGCTCAAGGTCAGCTA 175

QY 1330 ATTTCTCTCAAGAGATAATPCCACTGATTCAAAACAGCGCTTCTTAAATTAAATTCACAAAG 1389

DB 176 CTTTACTGAGGGAACAACAGAGCTCTGCTGGTGTCTCTACGAAAAACCTTCACTCTTTAAAG 235

QY 1390 AGGTGGAAAATGCTTTAAGTTCCCTATCAATACCGGTTTAGACGATAAACACTAGAC 1449

DB 236 AGGTGGAAACATGCTCTGGAACCGTCCCTTCTACACCAATTTGGACCGCTACACCATAGGT 295

QY 1450 GAAATATACAGCTTTTACAATGTAGACAATAACAAGAAATCTTGAAACCTACATATCACTCAT 1509

DB 296 GGAACATCGAAATTTCAATATTATAGACAGACATGCTAGACACCATCTTGTCAA 355

QY 1510 CAATATATAGTAAACACTGATTACCTAGTTGGCTGTGTGAAGATTTCTACACCTGCCAAT 1569

DB 356 ATCAACATACCAGTAGAGATATTTCTAGCGTTGAGTATTAGAGACTTCAGTTCTCTCAGT 415

QY 1570 CTATTTATCGTGAAGAAATTAAGAGTCTTTGAAAGCTGGGTGGTAGAGAAATAAGTTGGAAC 1629

DB 416 TTACTTACAGCAAGACTTCAACATCTTGAAGACTGGGTGAAGAGCTGAGGTTAGACC 475

QY 1630 AGCTCAAGTTTCTAGGCAAAAGACCGCTACTGTTATTTCTCTGTTGCTGCAACATTT 1689

DB 476 AGCTACAATTTGGCGACAGAAAGTTGGCATCTCTACTTCTGCTGCTGGCACCATGT 535

QY 1690 CGTCTCCGAATTATCAGATGGCGTATTTTCATGGCCCAAAATGGCATATTAACTACAG 1749

DB 536 TCTCTCTGAGTGCTGATGCTCGAACTTTTGGGGCCAAAATGGGTGCTGCTCAACTA 595

QY 1750 TAGTTGATGACTTTTCTTGATATPCGTTGGTGTACAATCGATGAATTGACCAACTGATTCAT 1809

QY	2408	CGGTGTCGAATTTT	TATACGCAAAACGATGACGGGTTTA	2444
Db	3336	GATGTGTCATGTTCT	ACTCTCGGACTGATGGGTTTA	3373
RESULT 15				
US-10-425-114-729				
; Sequence 729, Application US/10425114				
; Publication No. US2004003488A1				
; GENERAL INFORMATION:				
; APPLICANT: Lib, Jingdong				
; APPLICANT: Zhou Yihua				
; APPLICANT: Kovalic, David K.				
; APPLICANT: Screen, Steven E				
; APPLICANT: Tabaska, Jack E				
; APPLICANT: Cao, Yongwei				
; TITLE OF INVENTION: Nucleic Acid Molecules and Other M				
; TITLE OF INVENTION: Plants and Uses Thereof for Plant				
; FILE REFERENCE: 38-21(53313)B				
; CURRENT APPLICATION NUMBER: US/10/425,114				
; CURRENT FILING DATE: 2003-04-28				
; NUMBER OF SEQ ID NOS: 73128				
; SEQ ID NO 729				
; LENGTH: 1150				
; TYPE: DNA				
; ORGANISM: Zea mays				
; FEATURE:				
; OTHER INFORMATION: Clone ID: 7000808295_FLI				
US-10-425-114-729				
Query Match 10.3%; Score 288.8; DB 16; Le				
Best Local Similarity 59.1%; Pred. NO. 2.1e-57;				
Matches 513; Conservative 0; Mismatches 352; Inc				
QY	1578	CGTGAAGAATTA	AAAGGCTCTTCAAGGTCGGGTGAGAGATAA	1578
Db	3	CGGAGTGA	ACTTCGGCATCTTGATAGTTGGGTGAAGGAGAACAA	3
QY	1638	TTTGCTAGGCAAA	AGACCGCCTACTGTATTTCCTGTTGCTGCTG	1638
Db	63	TTTGCTGGCAG	AAACTCACAATTTGCTGCTGCTGCTGCTGCTG	63
QY	1698	GAATTATCAGAT	CGCGGTATTCATGGGCCAAAAAATGGCATATT	1698
Db	123	GAATTGCTG	TGACGCTCGCATTTTCATGGCCAAAAAATGGTGTCT	123
QY	1758	GACTTTTTTGAT	TCGGTGGTACAACTCGATCGAATGACCAACCT	1758
Db	183	GACTTCTTCG	ATGTTGGTGGATCAAAAGAGAAATTAGAAAACCT	183
QY	1818	AAATGGAAT	GTAGATGTCGCAAGGATTTGTTTCAGAGCATGT	1818
Db	243	AAATGGC	ATGGCCACCATGCACTGAGTTCTATTTCGGAAACAGGT	243
QY	1878	GCAATTAAAG	ATGCAATCTGTGTGATTTGGAGATGAAGCTTTTAA	1878
Db	303	GCTATTTAT	CAACAGTGAACCATCTTGGAGCAATGSCITTCG	303
QY	1938	GTAAC	TGACCATGTTATTCAAACTTGTTGGACTAATGAATAG	1938
Db	363	CTTACAA	ACCACCTAGTAGAATAATGGCTGGATTTGTTAAGATC	363
QY	1998	ATATGG	CAACAGATGCTTATGTGCCAACATTAATGAATATAT	1998
Db	423	GAATGG	CAGAGATGCGCAATATGTACCAACAGTTGAAGATACAT	423
QY	2058	TCATTTG	CAATAGCCCGAATGTCAAGCCGGCTATTTTACTTTGCT	2058
Db	483	TCATTTG	CACCTGGGCCAAATTTGTCTCCAGCATTTGATTTTGT	483
QY	2118	GAGG	GAATTTGAAAGCTCTGAATATCATATCTATTTAAGCT	2118
Db	543	GAGCATG	CTGTCGAAGATGAAGAGTACGATTAATTTATTTAGGCT	543

Db	236	AACTAAABACTGGAGCTTTGCACTTTGAAGAAAACAAAAGAAATTAAGAAATTTGTTGCA	295
Qy	335	AAATGTAGAAATTTCTGTTTCTTCAATATGACACAGCATGGGTAGCCATGGTCCCTCTCC	394
Db	296	CAAGGTTGAACCTTTTCAGTTTCTGCAATATGATCACTCATGGGTGGCAATGGTTCTCTCC	355
Qy	395	AAACATCACCAATCGCTTGTTCCTTGAGTGTCTCAATTTGGTTAAATTAATAACAGCT	454
Db	356	AAACTCTCTCAACCAACCTCTTTTCCCGAGTGTATAACTGGGTATAGATAGTCAACA	415
Qy	455	TAATGATGGTTCAATGGGTCTTTGTTAAATCACACTCAATATCAATCAATCAACCGCTTAA	514
Db	416	TGCTGATGGCTCATGGGCTACTCCACAAC-----GATCAGTTGCTGATGAA	463
Qy	515	AGATCTCTATCTTCAACATTAGCATGTATGTTGCATTAATAAGATGGAATTTGGGGA	574
Db	464	GGCCAATCTCTTATCTACATTAAGCATGTGTTCTTACTCTTAAGCGGTGGAATTAITGGGCA	523
Qy	575	AGATCAAAATAAATAAGGTCTAAGTTTATTCAGTCAAAATCTTGCTTCAGCTACTGAAAA	634
Db	524	TGATCATATGAGCAAGGCCCTTGATTTATCAAGTCTAATATAGCTTCAGCTACTGATGA	583
Qy	635	AAAGTCAACCATCTCCATTTGTTTGAATCATATTTCTGTTGCTTGGTTGCTTGAATGCGAA	694
Db	584	GAAACAAAGCTTCTCGGTGGGATTTGACATTAATTTTCCCTGGCATGATGAGTATGCTAA	643
Qy	695	AAACTTGGACATAAACCCTCTTCAAAACAAACAGATTTTGTGATGCTACATAAGAG	754
Db	644	AGACTTGAATTTGAATCTACCTCTGGCACCGAAGCGTGGATGCTTGGTTGGAAGAA	703
Qy	755	GGAATTTGGAGCAAAAA-----AGATGCCATTCMAATGAGATGGATGATCTTGGCGTA	808
Db	704	AGAGTTGGAGCTGAGAGCTGCAGAGCAACTCTGAAGGTGGAAGCCATTTAGCGTA	763
Qy	809	TATCTCTGAAGGACTCGGTAATTTATGATTTGGAATATGGTGAAGAAATATCAGATGAA	868
Db	764	TGTTTCAGAAGGAATTTGAAGAGTTACAGGACTGGGATATGGTCATGCAATATCAAAAGAA	823
Qy	869	AAATGTTCTGTTTCAACTCACCATCAGCAAGCTGCTGCTTCAATTAATCATCAAAA	928
Db	824	GAAATGATCACTGTTTAAATCTCCATCCACTACGGCAGCGGCTTTATGCAATAGAATGA	883
Qy	929	TCTGTTGCTGCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	988
Db	884	TGATGCTGTTTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	943
Qy	989	AGTTATCTCATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1048
Db	944	AATATATCTCTTGATATATATGCTCGAATACACATGGTTGATAGCCTTCAAAAATTCGG	1003
Qy	1049	AAATTTACACCAATTTTCAGAGTGAATTAATAATTAATTAATTAATTAATTAATTAATTA	1108
Db	1004	AAATTTCTCGCATTTCAAGAGAGATTAAGACGATTAATAGATGAACCTTACAGGTGTTG	1063
Qy	1109	GGTGAACAGAGATGAGCAAAATTTCAATGATGTTGTAACATGTGCTTTAGCCCTTTCCGTT	1168
Db	1064	GATGCAAGGAGAGAAAATATATTTAGATGCTTCBACTTTGTCGAATGSCCTTTTGAAT	1123
Qy	1169	ATTAAGATCAATGGGTATGAAGTTTCCCGAGATCCATTTGGCTGGAATTAATTAATAA	1222
Db	1124	GTTACGTTTGAAGGATATGATGTTTCTTCAGACCAAGTTGACTCAATTTTCAGAGATAT	1183
Qy	1223	-----TGAAATAGCTTTGAAGAGCAATATGACGCTCTTCAAAACATATCA	1267
Db	1184	CTTTCCCAATGGCTTTGGAGATATTTAAAGACTTCGGTGCCTCGCTGGAGTTATATAA	1243
Qy	1268	TGCGTCAATATATATTAACAGAGGATTTATCTTCT-----GGAAAAAATACTTGAAGTC	1324
Db	1244	GGCTCTCAGATTAATCACGCACCCCGATGAATCTGTTCTGGAATAATAAATCTTGGAC	1303
Qy	1325	AGCTGATTTCTCAAGAGATATATATCCATGATTAACAGGCTTTCTAAA-----	1376
Db	1304	TAGTCGTTTCTGGAAGCATGGATTAATCTAGTGATTTAGTTTGGTCTGATAGAACCGATAG	1363
Qy	1377	TTAATTTCAAAAGAGGTGGAAAAATGCTTCTTAAGTTCCCTATCAATAACGGTTTAGAAG	1435
Db	1364	TGTTCTTAAACBAGAGGCTGTAAATGCTTCTGAGTTCCCTATAATCAACTCTAGAAG	1423
Qy	1436	CATAAACCTAGACGAAATATACAGCTTTAATGATGACATACAGAAATTTCTGAAAC	1495
Db	1424	CCTAATAAGTAAGAGGCAATGGAAAGTTACAGTGGAGACATTTGTGAGGATTTCAAAATC	1483
Qy	1496	TACATATCACTCATCAAAATATAGTAACACATGATTACCTTAAGTTGGCTTGTGAAGATTT	1555
Db	1484	GCCATATGCTGCTTAAATTTTGGCCATCAAGATTTCTGGAACTTGTCTGATAGGATTT	1543
Qy	1556	CTACACCTGCCAATCTATTTATCTGTGAAGAAATTAAGAGTCTTGAAGGTGGGTGGTAGA	1615
Db	1544	CAATACCTCGCAACCCATTCATTTAAAGAACTCGAAGAGCTTCAAGATGGTGGTGA	1603
Qy	1616	GAATAAGTTGGACAGCTCAAGTTTGTAGGAAAAGACCGCTACTGTTATTTCTCTGT	1675
Db	1604	AAAAAATTTGGACGAGTTGAAATTTTTCAGACTGCACCTAGGTTACTGCTATTTTGTCTG	1663
Qy	1676	TGCTGCAACACTTCTGCTCCGCAATTTATCAGATGCGCTATTTATGCGGCCAAAAATCG	1735
Db	1664	GGCAGCGCCCTTACTGATCCCTGAATTCATGCTCATGCTGCTCATAGCATGGGCACAAATGG	1723
Qy	1736	CATATTAACTACAGTAGTTGATGACTTTTTCATATCGGTGGTACAAATCGATGAATGAC	1795
Db	1724	TGTGCTCAGACCGTGGTTGATGATTTCTATGATGGTGGAGGATCTGAAGAGGAAATGGA	1783
Qy	1796	CAACCTGATTAATGTTTGAATAATGGAATGATGTCGACAGGATTTGTTTTCAGA	1855
Db	1784	TAACTTTATAGAATTTGGTGGAAAAGTGGGATCCTGATGGGGAAGTGGTTTACTGTTTCCA	1843
Qy	1856	GCATGTTCCGATTTTATTTTATAGCATTTAAAGATGCAATCTGTTGGATTGGAGATGAAGC	1915
Db	1844	GGACGTTGAGATTTGATTTTCTTGACTGCACAGCAAGTTTGTGAATAGGAAGAGAGAGC	1903
Qy	1916	TTTTAAATGGCAAGCGCGATGTAACCTAGCCATGTTTATTCAAAATTTGTTGGTGAACATA	1975
Db	1904	TTTAGTATGCAAGGACGAGTGTATGAGGAATGTTATCGATGGTGGTCTGCTGCT	1963
Qy	1976	GAATGATGTTGAGAGAGCTATATGGAACAGAGATGCTTATGTCGCAACATAATTAATGA	2035
Db	1964	GAAGGTGATGAGAAAGAGAGCTGAATGGTTCGACAAATAAGGTAGTGCCTCAATGGGTGA	2023
Qy	2036	ATATATGAAAAACGCTTACGTGTCATTTGCTTTAGCCCGATGTCGAAGCCGCTATTTA	2095
Db	2024	ATATATGGAACAGCCCATGATCATTCGGTTGGACCTATAATCTTCCATGCTCTT	2083
Qy	2096	CTTTGTGGGCCCCAAATTAATCAGAGAGATGTTGAAAGCTCTGAATATCATTAATCTATT	2155
Db	2084	CTTTGTTGGACCTTAACTCTCAGAGGAATGATTGGAAGCTGTGAATACCAAGATTATA	2143
Qy	2156	TAACTAATGAGCACGCGAGCTGACTTCTTAACGATATCCATAGCTTCAAGAGGGAATTT	2215
Db	2144	TAACTGATGAGCACTGCTGGTCCCTTAAGATGATATTCGATCTTACGATGAGAAATG	2203
Qy	2216	TAAAGAGGCAAAATTAACCGCGTAGCATTTGCAATTTGAGTAAACGGAAGAGTGGGAAAGT	2275
Db	2204	CAAGAGGGAAGCTGATATTTCTGCTCTGTTGGATGATTTGATGGCGGTGATTAATGTCAC	2263
Qy	2276	GGAAGAGAGGTTTGTGGAGAGATGATGATGATTAATAAAACAGAGGAAGAAATTAAT	2335
Db	2264	CAAGAGGAGGCCATTTGAAGCAATTAAGGGGATTTTGAAGGGCGATGAAGAGCTGCT	2323
Qy	2336	GAAATTAATTTTGAAGAAATGATGATGATTTCTCTAGAGCTTGTAAAGATGCATTTTG	2395
Db	2324	GGGTTAGTTTTCAGAGAGAACACTCAA---TTCCAAGAGCTTGAAGATTTGTTCTG	2380
Qy	2396	GAACATGTCACGTGTTGAAATTTTTCAGCAAAACGATGACGGGTTTCTCGAAACAC	2455
Db	2381	GAAATGATGCTCATTTCTGATCTATTTTACATGGAAGATGATGGGTACACTTCAAATAG	2440

QY 2456 GATTCCTGATGACTGCAAGGACATCATTTACACCCGTTGGTCTTG 2502
 Db 2441 GTTGATGAACACTGTAAAGCCATGTTTGAACAACCCATGGATCTGG 2487

RESULT 2

US-08-727-308-3
 ; Sequence 3, Application US/08727308
 ; Patent No. 6020176
 ; GENERAL INFORMATION:
 ; APPLICANT: Yuji KAMIYA et al.
 ; TITLE OF INVENTION: KAURENE SYNTHASE
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Wenderoth, Lind & Ponack
 ; STREET: 805 Fifteenth Street, N.W., #700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/727,308
 ; FILING DATE: October 8, 1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33,367
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-371-8850
 ; TELEFAX:
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2658 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; US-08-727-308-3

Query Match 28.0%; Score 782.2; DB 3; Length 2658;
 Best Local Similarity 61.3%; Pred. No. 1.1e-191;
 Matches 1390; Conservative 0; Mismatches 823; Indels 54; Gaps 6;

QY 275 AACTAATCCCACTAATCTGATCATTTGATACACCAAGAACGATCCAAAACAGTTTAA 334
 Db 236 AACTAAAACCTGGAGCTTTGCACTTTGAAGAAACAAAAGAAAGAAATTAAGAAATTTGTCGA 295
 QY 335 AAATGTAGAATTTCTGTTTCTTCATATGACACAGCATGGTAGCCATGGTCCCTTCTCC 394
 Db 296 CAGGTTGAACCTTCAGTTTCTGATATGATGCAATGGTGGCAATGGTTCCTCTCC 355
 QY 395 AAACCTACCAAAATCGCCTTGTTTCCCTGAGTGTCTCAATTTGGTTAATTAATATCAGCT 454
 Db 356 AAACCTCTCAACCAACCTCTTTTCCCGAGTGTATAAAGCTGGTATAGTAGTCAACA 415
 QY 455 TAATGATGTTTCATGGGTCTTGTTAATACACTCAATAATCAATCAACCCGTTGCTTAA 514
 Db 416 TGCTGATGGCTCATGGGGCTTACTCCCAAC-----GATCAGTTGCTGATGAA 463
 QY 515 AGATTCTCTATCTTCAACATAGCATGTATGTTGCAATTAAGAGATGGATGTTGGGA 574
 Db 464 GGCCAAATCTCTTATCTACTATAGCATGTGTTCTTACTCTTAAGCGGTGGAATATTGGCA 523

QY 575 AGATCAATAAATAAGGTCTAAGTTTATGAGTCABATCTTGTCTGAGCTACTGAAA 634
 Db 524 TGATCATATGAGCAAGGCCCTTGATTTTATCAAGTCTAATATAGCTTACGTACTGATGA 583
 QY 635 AAGTCAACCATCTCCCATTTGGTTTGGACATCATATTTCTCGTTTGTCTGAGTATGCGAA 694
 Db 584 GAAACCAAGCTTCTCGGTGGGATTTGACATTAATTTCCCTGGCATGATTTAGTAGTCTAA 643
 QY 695 AAACCTGGACATAAACCTCTCTTTCAAACCAACAGATTTTGTGATGCTCATAGAAG 754
 Db 644 AGACTTGAATTTGAATCTACCTTGGCACCGACGACGTCGTGATGCTTGGTTTCGAAAGAA 703
 QY 755 GGAATTTGAGCAAAA-----AGATGCCATTTCAATGAGATGGATGATGATCTTGGCGTA 808
 Db 704 AGATTGGAGCTGAGAAGCTGCAGAAAGCAACTCTGAAGTGGAAAGGCTTATTTAGCGTA 763
 QY 809 TATCTCTGAAGGACTCGGTAATTTATATGATTTGAATATGTTGAAGAAATATCATAGTAA 868
 Db 764 TGTTCAGAAGGAATTTGAAAGTTACAGGACTGGGATATGTCATGCAATATCAAGAA 823
 QY 869 AAATGGTCTCTGTTTCAACTCAATCAGCAACAGCTCTGCTTTTCAATATCATATCAAAA 928
 Db 824 GAATGGATCACTGTTTAAATTTCCATCCACTACCGCAGCGCTTTTATGCAATAGAAATGA 883
 QY 929 TCCTGGTTGCTTAATTAATTTAAATTTCACTTTTGGCAAGTTTGTATGCTGAGTCCCAAC 988
 Db 884 TGAATGGCTGTTTGAATTAATTTTCTGCTCACTCTTACAAAAGTTTGTATGCTCAGTCCAC 943
 QY 989 AGTTTATCTCATGATTTATTTATCTCCGACTTTTCTATGTTTGACACAAATGAAAGATTAGG 1048
 Db 944 AATATATCTCTTGATATATATGCTCGATTACATGTTTATGATGCTTCAAAATTCGG 1003
 QY 1049 AATTTACACCATTTGAGAGTGGAAATTAATAATTTTATAGTGAACAAATCAGATGTTG 1108
 Db 1004 AATTTGCTGGCAATTTCAAGAGGAGATTTAGAAGCGTATTAGATGAAACTTACAGGTGTTG 1063
 QY 1109 GGTGGAACGAGATGAGCAATATTTCAATGAGTTTCTAACATGCTGCTTTAGCTTTTGGTT 1168
 Db 1064 GATCGAAGGAGGAAATATTTCTTAGATGCTTCACTTTGTGCAATGGCTTTCGAAAT 1123
 QY 1169 ATTAAGGATCAATGGGTATGAAGTTTCCAGATCCATTTGGCTGAAATTTACTAA----- 1222
 Db 1124 GTTACGTTTGAAGGATATGATGTTTCTTACAGCAGTGTGACTCAATTTTACAGAGATAT 1183
 QY 1223 -----TGAATTAGCTTTGAAAGACGAATATGACAGCTCTTTGAAAACATATCA 1267
 Db 1184 CTTTCCCAATTTGCTTGGAGGATATTTAAAGACTTCGGTCCCTCGCTGGAGTTATATAA 1243
 QY 1268 TGCGTCAATATATATACCAAGAGGATTTATCTTCT-----GGAAAACAAATCTTGAAGTC 1324
 Db 1244 GGCTCTCAGATTTACGCCACCCCGATGAATCTGTTCTGGAAAATATAAACTCTTGGAC 1303
 QY 1325 AGCTGATTTCTCAAAGAGATAATATCCATCTGATTCAAACAGGCTTTCTAAA----- 1376
 Db 1304 TAGTCGTTTCTGAAGCATGGAATTTCTAGTATTTAGTTTGGTCTGATAGAACCGTAG 1363
 QY 1377 -TTAATTCACAAAGAGTGGAAATGCTTTAAGTTCCTTATCAATACCGTTTTAGAACG 1435
 Db 1364 TGTGTTTAAACAGAGGCTGTTAATGCTTTAGTTCCTTGGTTCCTTATATGCAACTTAGAACG 1423
 QY 1436 CATAAACACTAGAGAAATATACAGCTTTTACATGTAGACAAATACAGAAATTTCTGAAAC 1495
 Db 1424 CCTAATAGTAAGGGCAATGGAAGTTACAGTGGAGACATTGTGAGGATTTCCAAATC 1483
 QY 1496 TACATATCACTCATCAATATTAAGTACACTGATTACCTAAGTTGGCTGTTGAAGATTT 1555
 Db 1484 GCCATATGCTCTGTTAAATTTTGCCCATCAAGATTTTCTGGAACCTGCTGTAGAGGATTT 1543
 QY 1556 CTACACCTGCAATCTATTATTCGTGAAGAAATTAAGAGTCTTTGAAAGTGGGTGGTAGA 1615
 Db 1544 CAATACCTGCAACGGCAATTCATCTTAAAGAACTTGAAGAGCTTCAAGAGATGGGTGGTTGA 1603
 QY 1616 GAATAAGTTGGACCGACTCAAGTTTGTAGGCAAAAGACCGCTACTGTTATTCTCTGT 1675

1604 AAACAAATGACGAGTGAATTTTTCAGACTGCACCTAGGTAATGCTATTTTGGCTGC 1663
1676 TGCTCCAAACATTTTCTGCTCCCGAATTAATCAGATCGGCGTATTTTCATGGCCAAAATGG 1735
1664 GCGCGGACCCCTTACTGATCTCTGAACTTTCATGCTCGCATAGATGGCGCAAAAATGG 1723
1736 CATATTAACACTAGTAGTGAAGCTTTTTCATATCGGTGTACAAATCGATGAATTCAC 1795
1724 TGCTCTCAGGACCGGTGGTGGATTTCTATGATGGTGAGGATCTGAAGAGGAATTCGA 1783
1796 CAACCTGATTTCAATGTTTGAATAATGGAATAGATTCGACAAAGGAATGTTGTTTCAGA 1855
1784 TAACCTTATAGAAATGGTGAAAAGTGGATCCTGATGGGAAGTGGGTTACTGTTCCAA 1843
1856 GAGTGTTCGGAATTTATTTTATAGCAATAAAGATGCAATCTCTTGGATGGAGATGAGC 1915
1844 GGAAGTTCGAGATGATTTCTTTCGACTGCACAGCAGTTTGTGAATAGGAAGAGAGC 1903
1916 TTTTAAATGGCAAGCGCGATGTAACCTAGCATGTTATTCAACTTTGGTTCGAACTAAT 1975
1904 TTTAGTATGCCAAGGACGAGTGTATAGGATGTTATGAGGATGTTATCGATGGTGGCTGCT 1963
1976 GAATAGTATGTTGAGAGAGCTATATGGAACAGAGATGCTTATGTGCCAACATTAAGA 2035
1964 GAAGGTGATGAGAAAGGAAGCTGAATGGTTCGACAAATAAGGTAGTGCCATCAATGGGTGA 2023
2036 ATATATGAAAACGCTTACGTGTCATTTGCTATTTAGCTTAGCCCGATGTGCAAGCCGCTATTTA 2095
2024 ATATATGAAAACGCTTATGATGCTGATTCGCTGTTGGACCTAATATCTTTCCATGCTCTT 2083
2096 CTTTGTGGGCCCCAATATATCAGAGGAGATGTTGAAAGCTCTGAATATCAATATCTATT 2155
2084 CTTTGTGGACCTTAACTCTCAGAGGAATGATGGAAGCTGTGAATACAGAGTTATA 2143
2156 TAAGCTTAATGAGCAGCAGCGGTGCACTTCTTAAACGATATCCATAGCTTCAAGAGGAAT 2215
2144 TAAGCTGATGAGCAGCTGCTGCTTAAAGATGATATTCGATCTTACGATAGAGAAATG 2203
2216 TAAGAGGCAAAATTAACCGCGGTAGCATTGCAATTTAGTAACGGAAGATGGGAAAGT 2275
2204 CAAAGAGGAAAGCTGAATATTTCTGCTCTGTGGATGATGATGCGGTGGTGAATGTCAC 2263
2276 GGAAGAGAGGTGTTGGAGAGATGATGATGATGATTAATAAACAAGAGGAAGAAATTAAT 2335
2264 CAAAGAGGAGCCATTGAAGCAATTAAGGGGATTTTGAGAGGCGGATAGAGAGCTGCT 2323
2336 GAAATTAATTTTGAAGAAAATGGTAGCATTTGTTCTTAGAGCTTTGTAAGATGCAATTTTG 2395
2324 GGGGTTAGTTTTCAGGAGAACACTACAA---TTCCAAGAGCTTTGTAAGGATTTGTTCTG 2380
2396 GAACATGTGTCAGTGTGAAATTTTAAAGCAACGATGACGGGTTTACTGGAACAC 2455
2381 GAAATTAATGATGCTCAATTTGTAATCTATTTTACATGGAAGATGATGGGTACACTTCAATAG 2440
2456 GATTCCTGATCTGTAAGAGACATCATTTTCAACACCGGTTGGTGGCTTG 2502
2441 GTTGAATGACACTGTAAAGCCATGTTTGAACACCCATGGATCTGG 2487

RESULT 3

US-09-614-912-23
; Sequence 23, Application US/09614912
; Patent No. 6677502
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Orozco, Buddy
; APPLICANT: Miao, Gou-Hau
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Lee, Jian Ming
; APPLICANT: Sakai, Hajime
; APPLICANT: Weng, Zude

; APPLICANT: Caimi, Perry G
; APPLICANT: Anderson, Shawn
; TITLE OF INVENTION: Plant Metabolism Genes
; FILE REFERENCE: B81378 US NA
; CURRENT APPLICATION NUMBER: US/09/614,912
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/143,401
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,412
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/146,650
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 23
; LENGTH: 2403
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1049)
US-09-614-912-23
Query Match 16.2%; Score 453.2; DB 4; Length 2403;
Best Local Similarity 54.9%; Pred. No. 6.9e-107;
Matches 1119; Conservative 0; Mismatches 808; Indels 111; Gaps 7;
QY 513 AAGAGATCTCTATCTTCAACATTAGCATGTTTGTGCAATTAAGAGATGGAATGTTGG 572
Db 72 AAGAGTCTCTATCTTCAACATTAGCATGTTTGTGCAATTAAGAGATGGAATGTTGG 131
QY 573 GAAGATCAATAAATAAAGGCTTAAAGTTTATGAGTCAAAATCTTCTCAGTCACTGAA 632
Db 132 AGAGAGAACATTTGGAGAGGACTGCATTTTCATCGGAGGAAATTTCTCTGTGTATGGAC 191
QY 633 AAAAGTCAACATCTCCCATTTGTTGACATCATATTTCTTGGTGTGTTGTTGTTGTTG 692
Db 192 GAGCAGTTCACTTCTCTATAGGTTTCAACTTCACTTCTTCTGTTTGTCTTACCTCGC 251
QY 693 AAAAAGCTTGACATAAACCTCTTTCAAAACAAAACAGATTTTAGTTTGTGTTACATAAG 752
Db 252 ATTGATATGGTTTGAATTTCTGTGAACAAATTTGATGCTGTGTGCAATTTCTTACC 311
QY 753 AGGAAATGGAGCAAAAAGATGCC-----ATTCAAATGAGATGGATGATGTTG 803
Db 312 CGGAGATGGAATTTGAAAGGCTGCTGTGATAGTTCTTTTGGAAAGAAAGCATATATG 371
QY 804 GCGTATATCTCTGAAGGCTCGTAAATTTATATGTTGGAATATGTTGGAAGAAATATCAG 863
Db 372 GCTTTTATCCAGAGGATTCGGAATATGCTGGAATGGAATGGAATGGAATGGAATG 431
QY 864 ATGAAAATGTTGTTCTGTTTCAACTCACTCACTCACTCACTCACTCACTCACTCACT 923
Db 432 AGGAAGATGGATCATTTGTTGAGCCTCTTCCACACTGCTGTTGCAATTAATCCAAA 491
QY 924 CAAAATCTGTTGTTCTTAAATTTAAATTCATCTTTTGGACAAAGTTTGGTAATGAGTC 983
Db 492 TACAACGACCAAGCCCTTCAATACCTAAATTTGCTTGTCAATGAATTTGCGAGTCAGTA 551
QY 984 CCAAGATTTATCTCTCATGTTTATTTATCCGACTTTTCTGATGTTTCACTCACTCACT 1043
Db 552 CCAGCAATGATCTCTCAAGGTTCAATTTGAGGTTCAATTTGAGGCGCTTGAAAA 611
QY 1044 TTAGGAATTTTCAACCACTTTTCAAGATGGAATTAATAATTTTATGATGAAATACATAGA 1103
Db 612 ATGGGAATTTCTCAGCGCTTTTGTGAGTGAATAGAAAGCATCTCTGGACATGGATCAAT 671

QY 1104 TGTGGTGGAAACGAGATGAGCAAAATATTCATGGATGTTGTAACTGTCCTTTAGCCTTT 1163
Db 672 TGCTGTTACAGATGATGAGGAACTCATGATGGACATAGCAACATTTGCAATGCAATTT 731
QY 1164 CGGTTATTAAAGATCAATGGGTATCAAGTTTCCAGATCCATTTGGCTGAAATTAATTAAT 1223
Db 732 CCGCTTTTGGAGATGAATGGTTTACGATGTTTCCCTCAGATGAGCTGTCTCACGTTCTCTGGA 791
QY 1224 G-----AAATAGCTTTGAAAGACGAATATGTCAGCTCTTGAA 1259
Db 792 GCTTCCACTTTCATGATTCATCAAGGATATTTAAATGATACAAATCCCTACTGGAA 851
QY 1260 ACATATCATGCTGCA---TATATTAATCAAGAGGATTTATCTTCTGGAAAAAATAATC 1316
Db 852 TTGTCAAGAGACCTCAAAAGTCACCTTATCAGAAAAAGATCTGATCTTAGATCGCATAGGT 911
QY 1317 TTGAAGTCAGCTGATTTCTCTCAAGAGATATATCCAC----- 1354
Db 912 TCCTGGCTGGCAACTTATTAAGAGATAAGATGTGCTGTAGTAAAGGTCAAAAAGACTCGA 971
QY 1355 -----TGATTCAAAACAGGCTTTCTAAATTAATTCACA----- 1386
Db 972 TTTTGGAGAGATGCTGCAACAAAATTTAAATCTCAATTTTCACCTTGGAGGTTCACTT 1031
QY 1387 -----AAGAGTGGAAAATGCTTTAAAGTTCCCTATCAATACCGGT 1427
Db 1032 TATGCTGTGTTTGTGNTTTCAGATCGAGTATGCTGTTAAATTTCCCTTGTATTCACA 1091
QY 1428 TTAGAAGCATAAACACTAGACGAATATACGCTTTTACATGTTAGACAATACAGAAATTT 1487
Db 1092 CTGGAGCTCTAGAACACAGAGAAACATCGAAATTTGATGCTGGGGTTCTCTGATG 1151
QY 1488 CTGAAAACATCATATCACTCATCAATATTAAGTAACTGATTAACCTAAAGTTGGCTGTT 1547
Db 1152 CT---AACAAACAAATCTCATCTTTCTGATCAATCAAGAAATTCCTAGCTTTGGCAGTC 1208
QY 1548 GAAGATTTTACACCTGCGCAATCTATTATTCGTGAAGAAATTAAGGTCCTTGAAGGTGG 1607
Db 1209 GAAGATTTTCACTTCTCTCAACGTTTACCGGATGAATTCGGCATCTTGATAGTTGG 1268
QY 1608 GTGTAGAGATAGTTGGACCGCTCAAGTTTGTAGGCAAAAGACCGCTACTGTTAT 1667
Db 1269 GTGAAGAGAACAGCTGGACCGCTCAATTTGCTGGCGAAGAACTGACATATTTGCTAT 1328
QY 1668 TTCTCTGTTGTGCAACACTTTCGCTCCGCAATATCAGATGCGCGTATTTTCATGCGCC 1727
Db 1329 CTGCTGCTGCTGCTACCGTATTTCTTCTGAATTTGCTGACGCTGCAATTTCAATGGCC 1388
QY 1728 AAAAATGGCATATTAACATACAGTATGTTGATGACTTTTTTGTATCGGTGGTACAAATCGAT 1787
Db 1389 AAAAATGGTGTCTTCAACACTGCTGTTGATGACTTCTTCGATGTTGGTGGATCAAAAGAA 1448
QY 1788 GAATGACCAACCTGATTCATGTTGTCAAAATGGAATGATGATGTCGACAAAGATTTGT 1847
Db 1449 GAATTGAAGAACCTGATAGACTAGTTTGAGAAATGGCATGGCCACCATGCAATTTGAGTTC 1508
QY 1848 TGTTTCAAGCATGTTTGGATTTTATTTTATAGCAATTAAGATGCAATCTGTTGGATTTGA 1907
Db 1509 TATTCGGAACAGGTGAAATAGTATTTTCTGCTATTTATACACAGTGAACCATCTTTGGA 1568
QY 1908 GATGAAGCTTTTAAATGGCAAGCGGATGTAATAGCCATGTTATTCAAACTTGGTTG 1967
Db 1569 GCAATGGCTTCTGAGCAGACAGGCGGTGATCTTCAAAACCATCTAGTAGAAATATGCTG 1628
QY 1968 GAACATAATAGTATGTTGAGAGAGCTATATGGAAGAGATGCTTATGTCGCAACA 2027
Db 1629 GATTTGTTAGATCTATGATGGTGGAGAGATGGCAGAGATGCCAATATGACCAACA 1688
QY 2028 TTAATGAATATATGAAAAACGCTTACGTGTCATTTGATTTAGGCCGATTTGCAAGCCG 2087
Db 1689 GTTGAAGAAATACATGACAAATGCTGTTGCTCATTTGACCTGGGCCCAATTTGCTCCCA 1748
QY 2088 GCTATTTACTTTTGTGGGGCCCAATATACAGAGAGATTTGTTGAAGCTCTGTAATATCAT 2147

RESULT 4

US-09-614-912-33

; Sequence 33, Application US/09614912

; Patent No. 6677502

; GENERAL INFORMATION:

; APPLICANT: Allen, Steve

; APPLICANT: Rafalski, Antoni

; APPLICANT: Orozco, Buddy

; APPLICANT: Miao, Gou-Hau

; APPLICANT: Famodu, Omolayo O.

; APPLICANT: Lee, Jian Ming

; APPLICANT: Sakai, Hajime

; APPLICANT: Weng, Zude

; APPLICANT: Caiji, Perry G

; APPLICANT: Anderson, Shawn

; TITLE OF INVENTION: Plant Metabolism Genes

; FILE REFERENCE: BB1378 US NA

; CURRENT APPLICATION NUMBER: US/09/614,912

; CURRENT FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: 60/143,401

; PRIOR FILING DATE: 1999-07-12

; PRIOR APPLICATION NUMBER: 60/143,412

; PRIOR FILING DATE: 1999-07-12

; PRIOR APPLICATION NUMBER: 60/146,650

; PRIOR FILING DATE: 1999-07-30

; PRIOR APPLICATION NUMBER: 60/170,906

; PRIOR FILING DATE: 1999-12-15

; PRIOR APPLICATION NUMBER: 60/172,959

; PRIOR FILING DATE: 1999-12-21

; PRIOR APPLICATION NUMBER: 60/172,946

; PRIOR FILING DATE: 1999-12-21

; NUMBER OF SEQ ID NOS: 204

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 33

; LENGTH: 1157

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (40)

; NAME/KEY: unsure

; LOCATION: (44)

; NAME/KEY: unsure

; LOCATION: (60)

; NAME/KEY: unsure

; LOCATION: (64)

; NAME/KEY: unsure

; LOCATION: (712)

; NAME/KEY: unsure

Db 1749 GCATTTGATTTTGTAGGCAAGAGCTATTAGAGCATGCTGTCAAAGATGAAGAGTACGAT 1808
QY 2148 AATCTATTAAAGTAAATGAGCAACGAGGTGAGCTTTTAAACGATATCCATAGCTTCAAG 2207
Db 1809 AAATTTATTAGCTAGTGAGCACTTGGCGGAGGCTCTCAATGACTACCAAAAGTTTAGAG 1868
QY 2208 AGGGAATTTTAAAGGAGGCAATTAACCGGCTAGCAATTTGCAATTTGAGTAAACGAGAAAGT 2267
Db 1869 AGGGAAGGCAACCGAGGGAAGCTGAATAGTGTCTTCTACTTGTCTCCACAGTGTGCT 1928
QY 2268 GCGAAAGTGGAAAGAGGTTGTGGAGGAGATGATGATGATGATTAATAAAACAAGAGGAAA 2327
Db 1929 TCTATGTCATAGAGCCGCTAAAGAGGCAATGAGAGTCCATAGAGCTGTCTAGGAGA 1988
QY 2328 GAATTAATGAATTAATTTTTTGAAGAAATGGTAGCATGTTCTTAGAGCTTTGTAAGAT 2387
Db 1989 GACTTGCTTAAGATTGGTTCT---CAGGAAAGAAAGTGTCTTCTTAGCCATGCAAGGAG 2045
QY 2388 GCATTTTGGAAACATGTGTCAGCTGTGTAATTTTTTTTACGCAACGATGACGGGTTTA 2445
Db 2046 CTCCTTGTGAAGATGTGTAAGATCTTCCACTGTTTTTACTCTCAGATGATGGATTTA 2103

```

; LOCATION: (721)
; NAME/KEY: unsure
; LOCATION: (723)
; NAME/KEY: unsure
; LOCATION: (732)
; NAME/KEY: unsure
; LOCATION: (738)
; NAME/KEY: unsure
; LOCATION: (757)
; NAME/KEY: unsure
; LOCATION: (763)
; NAME/KEY: unsure
; LOCATION: (782)
; NAME/KEY: unsure
; LOCATION: (809)
; NAME/KEY: unsure
; LOCATION: (813)
; NAME/KEY: unsure
; LOCATION: (824)
; NAME/KEY: unsure
; LOCATION: (826)..(827)
; NAME/KEY: unsure
; LOCATION: (849)..(850)
; NAME/KEY: unsure
; LOCATION: (858)
; NAME/KEY: unsure
; LOCATION: (863)
; NAME/KEY: unsure
; LOCATION: (872)
; NAME/KEY: unsure
; LOCATION: (874)
; NAME/KEY: unsure
; LOCATION: (888)
; NAME/KEY: unsure
; LOCATION: (930)..(931)
; NAME/KEY: unsure
; LOCATION: (938)
; NAME/KEY: unsure
; LOCATION: (942)
; NAME/KEY: unsure
; LOCATION: (961)
; NAME/KEY: unsure
; LOCATION: (967)
; NAME/KEY: unsure
; LOCATION: (971)
; NAME/KEY: unsure
; LOCATION: (981)
; NAME/KEY: unsure
; LOCATION: (988)..(989)
; NAME/KEY: unsure
; LOCATION: (993)
; NAME/KEY: unsure
; LOCATION: (999)
; NAME/KEY: unsure
; LOCATION: (1024)
; NAME/KEY: unsure
; LOCATION: (1028)
; NAME/KEY: unsure
; LOCATION: (1031)
; NAME/KEY: unsure
; LOCATION: (1035)
; NAME/KEY: unsure
; LOCATION: (1040)
; NAME/KEY: unsure
; LOCATION: (1049)..(1050)
; NAME/KEY: unsure
; LOCATION: (1068)
; NAME/KEY: unsure
; LOCATION: (1089)
; NAME/KEY: unsure
; LOCATION: (1105)..(1106)
; NAME/KEY: unsure
; LOCATION: (1121)

```

```

; NAME/KEY: unsure
; LOCATION: (1124)
; NAME/KEY: unsure
; LOCATION: (1131)
; NAME/KEY: unsure
; LOCATION: (1136)..(1137)
; NAME/KEY: unsure
; LOCATION: (1155)
US-09-614-912-33

```

```

Query Match      8.5%; Score 238.2; DB 4; Length 1157;
Best Local Similarity 58.1%; Pred. No. 1.3e-51;
Matches 433; Conservative 0; Mismatches 303; Indels 9; Gaps 1;

QY 505 CGTGTCTTAAAGATTCTCTATCTTCAACATTAGCATGTATTGTCATTGCAATTAAGAGATGGA 564
Db 54 CGGTCAACANGATGTTCTCTATCCACAGTTCGCAATGTTCTTCTGCGTGAAGAGATGGA 113
QY 565 ATGTTGGGGAAGATCAATAATAAAGGTCTAAGTTTATTGAGTCAAAATCTTGCTTCAG 624
Db 114 ATGTTGGCAGAGAGAAACATTTGGAGGAGACTGCATTTTCATCGGAGGAATTTCTCTGTG 173
QY 625 CTACTGAAAAGTCAACCATCTCCATGTTTGGACATCATATTTCTCTGTTTGTCTTG 684
Db 174 CTATGACGAGCAGTTCACTTCTCTTAGGTTTCACTTACCTTTCTCTGTTTGTCTTA 233
QY 685 AGTATGCGAAAACTTGGACATAAACCTCTTTCAAAAAACAAACAGATTTTATGTTGATG 744
Db 234 GCCTCGGCATTCATATGGTTTAGAATTTCTGTAAAGACAAATGATGCTCTGCGCATTC 293
QY 745 TACATAAGAGGAAATGGAGCAAAAAAG-----ATGCCATTCAATGAGATGGATG 795
Db 294 TTCACGCGCGGAGATGGAATTTGAAAAGGCTGGCTGTGGATAGTTCTTTTGGAGAAAAG 353
QY 796 GATACTTGGCGTATATCTCTGAAGGACTCGGTAATTTATATGATTGGAATATGTTCAAGA 855
Db 354 CATATATGGCTTTTATCCAGAGGATTCGGAATATGCTGACCTGGGATCAAGTATGA 413
QY 856 AATATCAGATGAAAAATGTTCTGTTTCAACTCACCATCAGCAACAGCTGCTGTTTCA 915
Db 414 AGTTTCAGAGGAAGATGGATCATGTTTCAGCACTCTTCCACAACTGCTGTTGCAATTA 473
QY 916 TTAATCATCAAAATCCTGGTGTCTTAATATTTAAATTCATTTTGGACAAGTTGGTA 975
Db 474 TCCACAAATACCAACGACCAAGCCCTTCAATACCTAAATTTGCTTGTCAATGAATTTGGCA 533
QY 976 ATGCGATCCCAACAGTTTATCTCTCATGATTTTATTTATCCGACTTTCTATGTTGACACAA 1035
Db 534 GTGCAATACCAATGATATCTTCAAGGTTACATGTCAGCTTCAATGGTGGACGCGC 593
QY 1036 TTGAAAGATTAGGAATTTCCACACCAATTTCCAGAGTGGAAATTAATAATGTTTATAGTAAA 1095
Db 594 TTGAAAAATGGGAATTTCTCAGCGCTTTGTGTCAGTGAATAAAAAACAATCCTCGACATGG 653
QY 1096 CATACAGATGTTGGTGGAAACGAGATGAGCAAAATATCATGATGTTTGAACATGCTT 1155
Db 654 CATACAAATGCTGTTTACAGAAAGATGAGAAATCATGATGAGCATAGCAACATTTGCNA 713
QY 1156 TAGCCCTTTCGTTTATTAAGGATCAATGGGTATGAACTTTCCCGAGATCCATTCGCTGAAA 1215
Db 714 TGGCATNCNCCCTTTTGANGATGANTGGTTACATGTTTCTCCTCNGATGANTGCTCAG 773
QY 1216 TTACTAATGAATTAGCTTTTGAAGA 1240
Db 774 TTGCTGANCCTTCCACTTTCCATGA 798

```

```

RESULT 5
US-09-398-395A-55
; Sequence 55, Application US/09398395A
; Patent No. 6468772
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph

```

; APPLICANT: No. 6468721, Joseph P.
 ; APPLICANT: Starks, Courtney M.
 ; APPLICANT: Manna, Kathleen R.
 ; TITLE OF INVENTION: SYNTHASES
 ; FILE REFERENCE: 07678-025001
 ; CURRENT APPLICATION NUMBER: US/09/398,395A
 ; CURRENT FILING DATE: 1999-09-17
 ; PRIOR APPLICATION NUMBER: 60/100,993
 ; PRIOR FILING DATE: 1998-09-18
 ; PRIOR APPLICATION NUMBER: 60/130,628
 ; PRIOR FILING DATE: 1999-04-22
 ; PRIOR APPLICATION NUMBER: 60/150,262
 ; PRIOR FILING DATE: 1999-08-23
 ; NUMBER OF SEQ ID NOS: 58
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 55
 ; LENGTH: 2861
 ; TYPE: DNA
 ; ORGANISM: Abies grandis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (3)...(2606)
 ; OTHER INFORMATION: abietadiene synthase
 US-09-398-395A-55

Query Match 7.0%; Score 194.8; DB 4; Length 2861;
 Best Local Similarity 47.8%; Pred. No. 3.1e-40;
 Matches 866; Conservative 0; Mismatches 85; Indels 87; Gaps 7;

QY	337	ATGTAGAAATTCGTCTTCATATGACACAGCATGGGTAGCCAGTCCCTTC	396
DB	397	ATGGGAAACGAAATCCCTCTGCATATGACACTGCTTGGGTAGCAGGATTC	456
QY	397	ACTCACCCAAATCGCTTGTTCCTGAGTGTCTCAATGTTGTAATTAATCAGCTTA	456
DB	457	ATGGCTCTGACAAACCTTCACCTTCTGAGAGCGTTTGAATGATTCCTCAAAATCAGTTGA	516
QY	457	ATGATGGTCTATGGGCTTGTGTAATCACTCATATCAATCAATCACCCTGCTTAAAG	516
DB	517	AAGATGGCTCTGGGCTGAAGGATTCATCTTC-----TGGCATTG	558
QY	517	ATTCCTCTATCTCAACATGATGATGTTGTAATTAAGATGGAATGTTGGGAG	576
DB	559	ACGAATCTGGCTACACTGCTGATGATATATACCTTACCTCTGGGCTACTGGGAGA	618
QY	577	ATCAATAAATAAAGGCTTAAGTTTATTTAGTCAATCTTCTTCACTACTGAAATA	636
DB	619	CACAAGTACAGAAGGATTAATCTTTCAGGACACAGCTGGAAAGATGAAG	678
QY	637	GTCACCATCT---CCCATGGTTTGAATCATATTTCTTCTGTTTCTGATGATGCGA	693
DB	679	CTGATAGTCATAGGCCAAGTGGATTTGAATAGTATTTCTTCTGCAATGCTAAAGGACTA	738
QY	694	AAACTTTGGACATAAACCTCTTCAAAACAAACAGATTTAGTTTGAAGCTACATAGA	753
DB	739	AAATCTTAGGCTGGATCTGCTTACGATTTGGCAATTCCTGAAACAAATCATCGAAAGC	798
QY	754	GGGAATTTGGAGCAAAAAGATGCCATTCAAATGAGATGGATG-----GATACTGG	804
DB	799	GGGAGGCTAGCTTAAAGGATTCCTGATGTTCTATGCTTATGCTTCCCAACACGTTAT	858
QY	805	CGTATATCTCTGAAGGCTCGGTAAATTTATATGATGGAATATGTTGGAAGAAATATCAGA	864
DB	859	TGATATCTTTGGAAGGTTTCAAGAAATAGTAGACTGGCAGAAATATGAATCTCAAT	918
QY	865	TGAAAATGTTCTGTTTCACTCAATCAGCAAGCAAGCTGCTGCTTCAATTAATCATC	924
DB	919	CCAGGATGATCATATTTCTAGCTCTCCGGCATCTACAGCGGCTGTATTCATGGGTACAG	978
QY	925	AAATCTGGTGTCTTAATTTAATTTCACTTTTGGACAGTTTGGTAATGCATCC	984
DB	979	GGACAAAAGTGTCTGGATTTCTTGAACTTTGTCTTGAAGAAATTCGAAACCATGTGC	1038

QY	985	CAACAGTTTATCCTCATGATTTATTTATCCGACTTCTATGTTGACACAAATGGAAGAT	1044
DB	1039	CTTGTCATATCCGCTTGTATCTATTTGAACTGTTTGTGGCGGTTGATACAGTTGAGCGGC	1098
QY	1045	TAGGAATTTCAACACATTTCCAGAGTGGAAATTAATAATGTTTTAGATGAAACATACAGAT	1104
DB	1099	TAGGTATCGATCGTCAATTTCAAAGAGGAGATCAAGGAAGCATTTGATTTATGTTTACAGCC	1158
QY	1105	GTGSGTGGAACGAG-----ATGAGCAAAATATTCATGATGATGTTGTAACAT	1149
DB	1159	ATTGGACGAAAGAGGCATTTGGATGGCGAGAGAGAACTCTGTTCTCTGATATGATGATA	1218
QY	1150	GTGCTTTAGCCCTTCGGTTATTAAGGATCAATGGGTATGAACTTTCCCGAGATCCATTTGG	1209
DB	1219	CAGCATGGGCTTCGAATCTTGATTTACATGGGATCAATGATATCTCTCAGATGTTTAA	1278
QY	1210	CTGAAATTTACTAATGAATTAGCTTTTGAAGAGCAATATGCAAGCTCTTGAACATATCATG	1269
DB	1279	AAACATTTAGAGATGAGAAATGGGAGTCTTTTGTCTTCTGGTCAAAACACAGAGGAG	1338
QY	1270	CGTACATATATTTATACC-----AAGAGATTTATCTTCTGAAACAAATCT	1317
DB	1339	TTACAGACATGTTAAACGTCATCTGTTTGCATGTTTCTATTTCCGGGAGAAACGATCA	1398
QY	1318	TGAAGTCAGCTGATTTCT-----CTCAAGAGATAATAT	1350
DB	1399	TGGAGAGCAAACTCTGTACCGAAAGGATCTGAGGAATGCTCTGGAATTTGGAGT	1458
QY	1351	CCACTGATTTCAACAGGCTTTCTAAATTTAAATTCACAAAGAGGTGGAAATGCTCTTAAGT	1410
DB	1459	CTTTTGACAAATGGCTTTTAAAGAAATATTCGGGAGAGGTAGAGTATGCATCAAT	1518
QY	1411	TCCCTATCATATCCGGTTTGAAGCGATATAACATAGACGAAATATACAGCTTTTACAAG	1470
DB	1519	ATCCCTGGCATAGAGTATGCCAAGTTGGAGGCTAGAACTATATTTGAAACTATGGC	1578
QY	1471	TAGACAATACAAAGATTTCTGAAACTACATCTCATCAATTAATTTAGTAACTGAT	1530
DB	1579	CAGATGATGTTGCTTGGAAATACTGTATATATGATGCCATACATTTCGAATGAAAGT	1638
QY	1531	ACCTAAGTTTGGCTGTTGAAGATTTTACACCTCCCAATCTATTTATCGTGAAGATTA	1590
DB	1639	ATTTAGAACTAGCGAACTGACCTTCAATAGGTGCACTATACACCAACACAGCTTC	1698
QY	1591	AAGCTCTGAAAGTGGTGGTAGAATAATGTTGGACCGAGCTCAAGTTTCTAGGCAAA	1650
DB	1699	AAGATCTGAAAGTGGTGGAAATCATCCGGTTTCAAGGATCTGAATTTCACTCGTAGC	1758
QY	1651	AGACCGCTACTGTTATTTCTCTGTTCTGCAACACATTTCTGCTCCCGAATTTACAGAT	1710
DB	1759	GTGTGACGGAATAATATTTCTCACCAGCATCTTATCTTTGAGCCCGAGTTTCTTAAGT	1818
QY	1711	CGGTATTTTATGGGCCAAAATGGCATATTAATACAGTAGTTTGTGATGATCTTTTTCGATA	1770
DB	1819	GCAGAGAGGTTTATACAAAATCTCCAAATTTCACTGTTATTTTAGATGATCTTTATGACG	1878
QY	1771	TCGGTCTCAATTCGATGATTTGACCAACCTGATTTCAATGTTTGAAGAAATGGAATGTAG	1830
DB	1879	CCCATGGATCTTTAGACGATCTTAAAGTTGTTTCAAGAATCAGTCAAAAGATGGATCTAT	1938
QY	1831	ATGTCGACAAAGGATTTGTTTTCAGAGCATGTTTCGAAATTTTATTTTATAGCATTTAAAGATG	1890
DB	1939	CACATAG---TGGACCAATGCCAACAAATGAATAATGTTTTTGTGGGTTTCTTACATA	1995
QY	1891	CAATCTGTTGGATGGAGATGAAGCTTTTAAATGGCAAGCGCGATGTAACATAGCCATG	1950
DB	1996	CTTTTAATGATATAGCAAAAAGAGGAGCTGAGAGGCAAGCGCGATGCTTAGGCTACA	2055
QY	1951	TTATTCAACTTGGTGGAACTAATGATAGTATGTTGAGAGAGCTTATATGACACAGAG	2010
DB	2056	TTCAAAATGTTTGGAAAGTCCAACTTTGAAGCTTTACAGAAAGAGAGAGATGCTCTGAAG	2115
QY	2011	ATGCTTATGTCACCAACATTAATGAATATATGAAACCGCTTACGTTGCTCATTTGCTATAG	2070

```

Db      2116 CTAATATGTGCGCATCTCTCAATGAATACATAGAGATGCGAGTGTGTCATAGCAATTGG 2175
QY      2071 GCCCGATTGT 2080
Db      2176 GAACAGTCGT 2185

RESULT 6
US-09-887-586A-55
; Sequence 55, Application US/09887586A
; Patent No. 6495354
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 64953541, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/887,586A
; PRIORITY FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 09/398,395
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 55
; LENGTH: 2861
; TYPE: DNA
; ORGANISM: Abies grandis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(2606)
; OTHER INFORMATION: abietadiene synthase
US-09-887-586A-55

Query Match      7.0%; Score 194.8; DB 4; Length 2861;
Best Local Similarity 47.8%; Pred. No. 3.1e-40;
Matches 866; Conservative 0; Mismatches 857; Indels 87; Gaps 7;

QY      337 ATCTAGAAATTTCTGTTTCTTCATATGACACAGATGGGTAGCGATGGTCCCTTCCTCCAA 396
Db      397 ATGGCGAAACGAATCCCTCTGCATATGACACTCTTGGGTAGCAAGATTCAGCAGTTG 456
QY      397 ACTCACCCAAATCGCCTTGTTCCTCGAGTGTCTCAATTGGTTAAATTAATTAATCACTTA 456
Db      457 ATGGCTCTGACACACCTCCTTCTCGAGACGGTTGAATGGATTCTTCAAAATCAGTTGA 516
QY      457 ATGATGGTTCATGGGTCCTTGTAAATCACTCATATCAATATCAATATCAATATCAATATCAAT 516
Db      517 AAGATGGGTCCTGGGGTGAAGATTCATCTTC-----TTGGCATATG 558
QY      517 ATTCTCTATCTCAACATTAAGCATGTATCTTTCATTTAAAGATGGAATGTTGGGGAAG 576
Db      559 ACAGATATCTGGCTACACTTTCGATATATTAATACCTTACCTCTGCGGTACTTGGGGAGA 618
QY      577 ATCAAAATAAATAAGGTCTAAGTTTATTCAGTCAATCTTGTCTTCACTTCACTTCACTTCACT 636
Db      619 CACAAGTACAGAAAGGTATTGAATCTTTCAGGACACAAGCTGGAAGATGGAAGATGAAG 678
QY      637 GTCAACCATCT---CCCATTTGGTTTTCACATCATATTTCTCTGTTTCTGTTGATGATGCGA 693
Db      679 CTGATAGTCATAGGCCCAAGTGGATTTGAAATAGTATTTCTGCAATGCTAAAGAGCTA 738
QY      694 AAAAATTTGGACATAAACTCTCTTTCATAAAACACAGATTTTATTTTATGCTCATATAAGA 753
Db      739 AAATCTTAGGCTTGATCTGCCTTACGATTTGCCATTCCTTGAACAACATCATCGAAAGC 798
QY      754 GGGAAATTTGGACAAAGATGCGCATTCAAATGAGATGGATG-----GATACCTTGG 804

```

```

Db      799 GGGAGGCTAAGCTTAAAGGATTTCCCACTGATGTTCTCTATGCGCCTTCCAAACAGCTTAT 858
QY      805 CGTATATCTCTGAAGGACTCGGTAAATTTATATGATTTGGAATATGATGTTGAAGAAATATCAGA 864
Db      859 TGTATTTCTTGGAAAGGTTTACAGAAATAGTAGACTGGCAGAAATATATGAACCTTCAAT 918
QY      865 TGAATAATGTTTCTGTTTTCAACTCACCATCAGCAACAGCTGCTGCTTTCATTAATCATC 924
Db      919 CCAAGGATGGATCATTTCTCAGCTCTCCGGCATCTACAGCGGCTGTATTCATGCGCTACAG 978
QY      925 AAAATCCTGGTGTCTTAAATTTAAATTTCACTTTTGGACAAGTTTGGTAAATGAGTCC 984
Db      979 GGAACAAAAAGTGTGTTGATTTCTTTGAACTTTTGTCTTGAAGAAATTCGGAACCCATGTC 1038
QY      985 CAACAGTTTATCTCATGATTTATTTATTCGGAATTTCTATGTTGACACAAATGAAAGAT 1044
Db      1039 CTTGTCACTATCCGCTTGAATCTATTTGAACGTTTGGCGGCTGATACAGTTGAGCGGC 1098
QY      1045 TAGGAATTTACACCATTTTCAGAGTGGAAATTTAAATTTTAAATTTTGTAGTAAACATCAGAT 1104
Db      1099 TAGGTATCGATCGTCAATTTCAAAGAGGAGATCAAGGAACATTTGGATTTATGTTTACGCC 1158
QY      1105 GTTGGTGGAAACGAG-----ATGACCAATATTTTCATGGATTTTGAACAT 1149
Db      1159 ATTGGGACGAAAGAGGCAATTGATGGCGAGAGAGAAATCTCTGTTCTGATATATGATATA 1218
QY      1150 GTGCTTTAGCCTTTTCGTTTATTAAGATCAATAGGTTATGAAGTTTCCCGAGATCCCATTTGG 1209
Db      1219 CAGCATATGGCCTTTGAACTTTGAGATTTAGATACATGATGATGATGATGATGATGATGAT 1278
QY      1210 CTGAATTTACTAATGAATTTAGCTTTTGAAGAGCAATATGCACTCTTGAACATATCATG 1269
Db      1279 AAACATTTAGAGATGAGATGGGAGTTCTTTGCTTCTTGGGTCAAACACAGAGAGGAG 1338
QY      1270 CGTCACATATATATAC-----AAGAGATTTATCTTCTCGAAACCAATCT 1317
Db      1339 TTACAGACATGTTAAACGCAATCGTTGTTGTTTCAATGTTTCAATTTCCGGGAGAAACATCA 1398
QY      1318 TGAAGTCAGTCATTTTC-----CTCAAAGAGATTAATAT 1350
Db      1399 TGGAGAAGCAAACTCTCTACCGAAGGTATCTGAGGAATGCTCTGGAATATGTTGGATG 1458
QY      1351 CCATGATTTCAAACAGGCTTTCTAAATTAATTTCAAAAGAGTGGAAATGCTCTTAAAGT 1410
Db      1459 CTTTTGACAAATGGGCTTTTAAAGAGATATTTGGGGAGAGGTAGATGTCATCAAT 1518
QY      1411 TCCTATCATACCGGTTTGAACGCATAAACAACATAGACGAATATACAGCTTTCAATG 1470
Db      1519 ATCCCTGGCATAGATATGCCAAGTTGGAGCTAGAAGCTATATATGATGCCATACATTTGAAAT 1578
QY      1471 TAGACAATACAGAAATTTCTGAAACATACATATCACTCATCAATATTTAGTAAACATGAT 1530
Db      1579 CAGATGATGTGGCTTTGGAAACCTGTATATATGATGCCATACATTTGCAATGAAAGT 1638
QY      1531 ACCTAAGTTGGCTGTTTGAAGATTTCTACCTGCGCAATCTATTTATTCGTGAGAAATTA 1590
Db      1639 ATTTAGAATCTAGCGAACTGGACTTCAATAAGGTGCAGTCTATACCAACAGAGAGCTTC 1698
QY      1591 AAGGCTTTGAAAGGTGGGTAGAGAATAAGTTGGACACAGCTCAAGTTTGTCTAGGCAAA 1650
Db      1699 AAGATCTCGAAGGTGGTGAATCATCCGGTTTCAGGATCTGAATTTCACTCTGAGC 1758
QY      1651 AGACCCGCTACTGTTATTTCTGTTGCTGCAACACATTTGCTCCGCAATTTACAGATG 1710
Db      1759 GTGTGACGGAATATATTTCTCAGCGCATCTTTATCTTTGAGCCGAGTTTCTTAAAGT 1818
QY      1711 CGGTTATTTCTGGGCCAAATATGGCATATTAATCTACAGTATGATGATGATGATGATGATGAT 1770
Db      1819 GCAGAGAGGTTTATACAAAACTTCCAAATTTCACTGTTATTTTAGATGATTTTATGAGC 1878
QY      1771 TCGGTGTTCAATCGATGAAATTCACCAACCTGATTTCAATGTTTGAAGAAATGGAATGTAG 1830
Db      1879 CCCATGATCTTTAGACGATCTTAAAGTTGTTTCACAGATCAGTCAAAAGATGGGATCTAT 1938

```

Qy		1831	ATGTCGACAAAGATTGTGGTTCCAGACATGTTCCGATTTTATTTTTTAGCATTTAAAGCATG	1890
Db		1939	CAC TAG --- TGGACCAATGCCACAACAATGAAAATAATGTTTGTGGGTTTCTCACATA	1995
Qy		1891	CAATCTGTTGGATTGGAGATGAAGCTTTTAAATGGCAAGCGCGATGTAACTAGCCATG	1950
Db		1996	CTTTTAATGATATAGCAAAAGAAGCAGCTGAGAGGCCAAGGCAGTGCTGCTAGGCTACA	2055
Qy		1951	TTATTCAAACCTTGCTTGGAACTAATGAATAGTATGTTGAGAGAAGCTATATGACACAAG	2010
Db		2056	TTCAAAATGTTTGGAAAGTCCAATCTGAAGCTTACACGAAGAAGCAGAATGCTCTGAAG	2115
Qy		2011	ATGCTTATGTGCCAACATTAATAATGAATATATGAAAAAGCGTTACGTGTCAATTTGCCATTAG	2070
Db		2116	CTAAATATGTGCCATCCTTCAATGAATACATAGAGAATGCGAGTGTGTCAATAGCATTTGG	2175
Qy		2071	GCCCCGATTGT	2080
Db		2176	GAACAGTCGT	2185

RESULT 7

```

US-09-895-752-55
; Sequence 55, Application US/09895752
; Patent No. 6559297
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 65592971, Joseph P.
; APPLICANT: Marks, Courtney M.
; APPLICANT: Mann, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/895,752
; CURRENT FILING DATE: 2001-06-29
; PRIORITY APPLICATION NUMBER: 09/398,395
; PRIORITY FILING DATE: 1999-09-17
; PRIORITY APPLICATION NUMBER: 60/100,993
; PRIORITY FILING DATE: 1998-09-18
; PRIORITY APPLICATION NUMBER: 60/130,628
; PRIORITY FILING DATE: 1999-04-22
; PRIORITY APPLICATION NUMBER: 60/150,262
; PRIORITY FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 55
; LENGTH: 2861
; TYPE: DNA
; ORGANISM: Abies grandis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(2606)
; OTHER INFORMATION: abietadiene synthase
US-09-895-752-55

```

Query Match	7.0%;	Score 194.8;	DB 4;	Length 2861;
Best Local Similarity	47.8%;	Pred. No. 3.1e-40;		
Matches 866;	Conservative 0;	Mismatches 857;	Indels 87;	Gaps 7;
QY	337	ATGTAGAAATTTCTGTTTCTTCATATACACAGCATGGGTAGCATGGTCCCTTCTCAA	396	
Db	397	ATGGCGAAGCAATCCCTCTGCATATACACTGCTTGGGTAGCAAGGATTTCCACAGTTG	456	
QY	397	ACTCACCAATCGCCTTGTTTCCCTCAGTGTCTCAATTGCTTAATTAATAATCAGCTTA	456	
Db	457	ATGGCTCTGACACCCCTCACTTTCTCGAGCGTTGAATGATTCTTCAAATCAGTTGA	516	
QY	457	ATGATGGTTCATGGGCTTGTTTAATCACACTCATATATCAATCAACCCGTGCTTAAAG	516	
Db	517	AAGATGGGTCTTGGGTTGAAGGATTCATTCTC-----TTGCGCATATG	558	
QY	517	ATTCTCTAICTTCAACATTAGCATGTATCTTGCATTAAAAAGATGGTGGGGAAG	576	

[illegible]

QY	1591	AAGTCTTTGAAAGGTGGGTGTAGAGATAAGTTGGACCAAGTCAAGTTTGTAGGCAA	1650
Db	1699	AAGATCTTCGAAAGGTGGTGAATCATCCGGTTTCAACGGATCTGAATTTCACTCGTGAGC	1758
QY	1651	AGACGGCTACTGTTATTTCTGTTGCTGCAACACTTTTCGTCCTCCCGAATTTATCAGATG	1710
Db	1759	GTGTGACGGAATATATTTCTCACCGGCATCTTTATCTTTGAGCCCGAGTTTCTAAGT	1818
QY	1711	CGCGTATTTCAATGGCCCAAAAANGCAATATTAACCTACAGTAGTTGATGACATTTTGTGATA	1770
Db	1819	GCAGAGAGGTTTATACAAAACCTTCCAAATTTCACTGTTTATTTTAGATGATCTTTATGACG	1878
QY	1771	TCGGTGGTACATCATGATGAATTCACCACTGATTCATGTTGTTGAAAATGGAATGTAG	1830
Db	1879	CCATGGATCTTTAGACGATCTTAAGTTGTTCCACAGATCAGTCAAAAGATGGGATCTAT	1938
QY	1831	ATGTCGACAAAGATTTGTTGTTCCAGAGCATTTCCGATTTTATTTTATAGCATTAAGAATG	1890
Db	1939	CACTAG---TGGACCAATGCCACAACAATGAAATATGTTTGTGGTTTCTACAATA	1995
QY	1891	CAATCTGTGTGAGATGAGCTTTTAAATGGCAAGCGCGCGATGTAACTAGCCATG	1950
Db	1996	CTTTTAAATGATATAGCAAAAGAGACGCTGAGAGGCAAGGCGCGATGTGTAGGCTACA	2055
QY	1951	TTATTCAAACTTGGTTGGAATCAATCAATAGTAGTATTGAGAGAAGCTATATGCAACAAG	2010
Db	2056	TTCAAAATGTTTGGAAAGTCCAACTTGAAGCTTTACAGAAAGAGCAAGATGCTCTGAAG	2115
QY	2011	ATGCTTATGTGCAACATTAATGAATATATGAAAACGCTTTACGTGTCAATTTGCATTTAG	2070
Db	2116	CTAATATGTGCCATCTTCAATGATACATAGAGATCGAGTGTGTCAATAGCATTTGG	2175
QY	2071	GCCCGATTTG 2080	
Db	2176	GAACAGTCTG 2185	
RESULT 8			
US-09-903-012B-55			
; Sequence 55, Application US/09903012B			
; Patent No. 6569656			
; GENERAL INFORMATION:			
; APPLICANT: Chappell, Joseph			
; APPLICANT: No. 65696561, Joseph P.			
; APPLICANT: Starks, Courtney M.			
; APPLICANT: Manna, Kathleen R.			
; TITLE OF INVENTION: SYNTHASES			
; FILE REFERENCE: 07678-025001			
; CURRENT APPLICATION NUMBER: US/09/903,012B			
; CURRENT FILING DATE: 2001-07-11			
; PRIOR APPLICATION NUMBER: 60/100,993			
; PRIOR FILING DATE: 1998-09-18			
; PRIOR APPLICATION NUMBER: 60/130,628			
; PRIOR FILING DATE: 1999-04-22			
; PRIOR APPLICATION NUMBER: 60/150,262			
; PRIOR FILING DATE: 1999-08-23			
; NUMBER OF SEQ ID NOS: 58			
; SOFTWARE: FastSeq for Windows Version 3.0			
; SEQ ID NO 55			
; LENGTH: 2861			
; TYPE: DNA			
; ORGANISM: Abies grandis			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (3)...(2606)			
; OTHER INFORMATION: abietadiene synthase			
US-09-903-012B-55			
Query Match 7.0%; Score 194.8; DB 4; Length 2861;			
Best Local Similarity 47.8%; Pred. No. 3.1e-40;			
Matches 866; Conservative 0; Mismatches 857; Indels 87; Gaps 7;			

Db 1459 CTTTGCACAAATGGGCTTTTAAAGAAATATTCGGGAGAGGTAGAGTATGCACTCAAAAT 1518
QY 1411 TCCCTATCAATACCGGTTTAGAACCAATAAACAACACTAGACGAAATATACAGCTTTTCAATG 1470
Db 1519 ATCCCTGGCATAAGAGTATGCCAAGGTTGGAGGCTAGAGCTATATTTGAAAACTATGGGC 1578
QY 1471 TAGACAAATACAGAAATCTGAAAACTACATATCACTCATCAATATTAATTAACACTGATT 1530
Db 1579 CAGATGATGTGGCTTGGAAAACTCTATATATATGATGCCATACATTTTCAATGAAAGT 1638
QY 1531 ACCTAAGTGTGGCTTTGAAAGTTTCACTACCTGCCAATCTATTTATCGTGAAGAAATTA 1590
Db 1639 ATTTAGAACTAGCGAAACTGGACCTTCAATAAGGTCAGTCTATACACCAACAGAGCTTC 1698
QY 1591 AAGGCTTTGAAAGTGGGTGGTAGAGAAATAGTGGACCCAGCTCAAGTTTGTAGGCAAA 1650
Db 1699 AAGATCTTGAAGTGGGTGGAAATCATCCGGTTTCAACGATCTGAATTTCACTCGTGAGC 1758
QY 1651 AGACCGCTACTGTTTATTTCTGTTGCTGCAACACTTTCGTCCTCCGAAATTTATCAGATG 1710
Db 1759 GTGTGACGAAATATATTTCTCACCGGATCTTTATCTTTGAGCCCGAGTTTCTAAGT 1818
QY 1711 CGGTAATTTGAGCCCAAAATAGGCATATTAACATACAGTATGATGACTTTTGTGATA 1770
Db 1819 GCAGAGAGGTTTATACAAAACTTCCAAATTTCACTGTTATTTTATGATGATCTTTATGAGC 1878
QY 1771 TCGGTGGTACAAATCGATGAATGACCAACCTGATTCATGTTGAAATAGTGGAAATGGAATG 1830
Db 1879 CCCATGATCTTTAGACATCTTAAAGTTTTCACAGATCAGTCAAAAGATGGGATCTAT 1938
QY 1831 ATGTCGCAAGGATGTTGTTTTCAGAGCATGTTGCGATTTTATTTTATGCAATTAAGATG 1890
Db 1939 CACTAG--TGGACCAATGCCACACAAATGAAATATGTTTGTGGGTTTCTACAAATA 1995
QY 1891 CAATCTGTGATGAGATGAAGCTTTTAAATGGCAAGCCGCGATGTAATAGCCCATG 1950
Db 1996 CTTTAAATGATATACAAAAGAGACGTCGAGAGCAAGGCGCGATGTCGTAGCTACA 2055
QY 1951 TTAATCAAACTGTTGGAATGATGATGATGATGTTGAGAGCACTATATGGAAGCAAG 2010
Db 2056 TTAATAATGTTTGAAGTCACTGAGCTTACAGAGCTTACAGAAAGAGCAGAGATGCTGAAG 2115
QY 2011 ATGCTTATGTCACAACTTAATTAATGAATATATGAAAAAGCTTACGTGTCATTTGCAATTAG 2070
Db 2116 CTAATATGTCCTCTTCAATGAATACATAGAGATGCGAGTGTCTCATAGCTTGG 2175
QY 2071 GCCGATGT 2080
Db 2176 GAACAGTCGT 2185

RESULT 9

US-09-900-797-55
; Sequence 55, Application US/09900797
; Patent No. 6645762
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 66457621, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/900,797
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/09/398,395
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 55
; LENGTH: 2861
; TYPE: DNA
; ORGANISM: Abies grandis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(2606)
; OTHER INFORMATION: abietadiene synthase
US-09-900-797-55

Query Match 7.0%; Score 194.8; DB 4; Length 2861;
Best Local Similarity 47.8%; Pred. No. 3.1e-40;
Matches 866; Conservative 0; Mismatches 857; Indels 87; Gaps 7;

QY 337 ATGTAGAAATTTCTGTTTCTTCAATAGACACAGCATGGTATGCCATGTCCCTTCTCCAA 396
Db 397 ATGGCGAAACGAATCCCTCTGCATATGACACTGCTTGGTATGCAAGGATTCACGAGTTG 456
QY 397 ACTCACCCAAATCGCTTGTTCCTCGAGTCTCAATGCTTAAATTAATCAAGCTTA 456
Db 457 ATGGCTCTGACCAACCTCACTTCTCGAGACGGTGAATGGAATCTTCAAAATCAGTTGA 516
QY 457 ATGATGGTTTCATGGGCTCTTCTTAATCACACTCATATAATCAATCAATCACCCTGTTGCTTAAAG 516
Db 517 AAGATGGGCTCTTGGGTGAAGGATTTCTATTC-----TTGGCATATG 558
QY 517 ATCTCTTATCTTCAACATAGCATGTATGTTGCAATTAAGATGGAATGTTGGGGAAG 576
Db 559 ACAGAACTACTGGCTACACTTGCATGTATATTACCTTTACCTCTGGCGTACTGGGAGA 618
QY 577 ATCAAAATAATAAAGGCTAAGTTTATTTAGTCAATCTTGTCTTCACTACTGAAAAA 636
Db 619 CACAGTACAGAAAGGATTTGAATTTCTCAGGACACAACTGGAAGATGAAGATGAAG 678
QY 637 GTCAACCATCT---CCCATTTGTTTTCACATCATATTCTCTGTTTGTCTCAGTATGCGA 693
Db 679 CTGATAGTCATAGCCCAAGTGGATTTGAATAGTATTTCTTGCATGCTTAAAGGAGCTA 738
QY 694 AAAAATTGGACATAAACCTCTTTCAAAACAAACAGATTTTGTATGCTACATAAGA 753
Db 739 AAATCTTTAGGCTGGATCTGCTTACGATTTGCCATTTCTGAAACAAATCATCGAAAAAGC 798
QY 754 GGGAAATGGAGCAAAAAGATGCCATTTCAATGAGATGATG-----GATCTTGG 804
Db 799 GGGAGGCTAAGCTTAAAGGATTTCCCACTGATGTTCTCTATGCTTCCAAACAGCTTAT 858
QY 805 CGTATATCTCTGAAGGACTCGGTAATTTATATGATGGAATATGTTGGAAGAAATATCAGA 864
Db 859 TGTATTTCTTTGGAAGGTTTACAAGAAATAGTAGACTGCGAGAAATATGAACTTCAAT 918
QY 865 TGAATAATGTTCTGTTTTCAACTCCATCAGCAACAGCTGCTGCTTTTCAATTAATCATC 924
Db 919 CCAAGGATGATCATTTCTCAGCTCTCCGCACTACAGCGGCTGTATTCTATGCGTACAG 978
QY 925 AAAATCCTGTTGTTCTTAATTTAAATTTCACTTTTGGACAGATTTGGTAATGCAATCC 984
Db 979 GGAACAAAAGTCTGGATTTCTTGAATTTGTCTTGAAGAAATTCGGAACCATGTGC 1038
QY 985 CAACAGTTTATCTCATGATTTATTTATCCGACTTTCTATGTTTGACACAAATTCGAAGAT 1044
Db 1039 CTTGTCACTATCCGTTGATCTATTTGAAGTTTGTGGCGGTTGATACAGTTGAGCGGC 1098
QY 1045 TAGGAATTTCAACCAATTTCAAGTGGAAATTAATAATGTTTGTAGTGAACATACAGAT 1104
Db 1099 TAGGTATCGATCGTCATTTCAAGAGAGGAGATCAAGGAAGCATTTGGATTATGTTTACAGCC 1158
QY 1105 GTTGGGTGGAACGAG-----ATCAGCAAAATTAATCAATGATGTTGTAACAT 1149
Db 1159 ATTGGGACGAAAGAGGCAATTTGGATGGCGGAGAGAGAAATCTCTTCTGATTTGATGATA 1218
QY 1150 GTGCTTTTAGCCTTTTCGGTTATTAAGGATCAATGGGTATGAAGTTTCCCAAGATCCATGG 1209
Db 1219 CAGCCATGGGCTTCGAATCTTGAGATTACATGATACATATGATCTCTCAGATGTTTAA 1278

Db 474 TTCTCTGATGATCTGAAGCCACGGTTCTCAGGCCCTCACTGGGTTTCAACA 533
 QY 449 TCAGCTTAATGATGTTTCATGGGCTTGTGTTAATCACACTCAATATCAATCAATCAACCGTT 508
 Db 534 CCAGCTCCAGGATGATCGTGGGTATCGAATCGACCTTTAGTTTATGCGATCGATTGCT 593
 QY 509 GCTTAAAGATCTCTATCTTCAACATTAGCATGTTGTTGATTAATAAAGATGAATGT 568
 Db 594 -----TAAACAGCAACCAATCTGTTATCGCCCTCTCGGTTTGGAAAC 635
 QY 569 TGGGAAGATCAATAAATAAAGGCTTAAGTTTATTTAGTCAAAATCTTCTGCTTCACTAC 628
 Db 636 AGGCACAGCCAAATGATACAAAGGCTGCTGAGTTTATGAGAGAACTAAGATTAATCAA 695
 QY 629 TGAATAAAGTCAACCAATCTCCATGTTGTTGATCAATCAATTTCTGTTGTTGTTGTTGTT 688
 Db 696 TGAGGAAGATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 752
 QY 689 TGGGAATACTTGGACATAAATCTCTTCAAAACAAACAGATTTTATGTTGATGTTGTTGTTGTT 748
 Db 753 GSCAAAGCGTTGGGATCAATCTTCTTACGATCTTCCATTTATCAATTTATTTGTTGTTGTT 812
 QY 749 TAAGAGGAATTTGGAGCAAAAGATGCCAT-----TCAATGAGATGATGATGATGAT 802
 Db 813 AACACGGAGAGCGAGCTTACAGATGTTCTGCGGAGAGACAAATATTCCAGCCAAAT 872
 QY 803 GCGGTATATCTCTGAAGGACTCGGTATTTATATGATGATTTGGAATATGGTGAAGAAATATCA 862
 Db 873 GTTGAATGCGTTGGAAGGCTCTGAGGAAGTTATTGATCTGGAACAAGATTTATGAGGTTTCA 932
 QY 863 GATGAATAATGTTCTGTTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 922
 Db 933 AAGTAAAGATGATGATCTTCTGAGTCTCCCTGCTCCATGCTGCTGCTGCTGCTGCTGCTGCT 992
 QY 923 TCAAAATCTGTTGTTCTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 982
 Db 993 AGGGAGCAAAATGTTTCACTTTTCACTTTTCACTTTTCACTTTTCACTTTTCACTTTTCACT 1052
 QY 983 CCAACAGTTTATCTCATGATTTATTTATCGAGCTTTTATTCGAGCTTTTATGTTGACACAAATGNAAG 1042
 Db 1053 GCCTGTATGATTTTCACTGATCTGCTGGAACGCTTTTCTGCTGTTGATTAACATTTGAGCA 1112
 QY 1043 ATTAGGAATTTCAACCAATTTCAAGTGGAAATTTAAATAATTTTAAATTTTAAATTTTAAAT 1102
 Db 1113 TCTCGGAATCGTCCGCTTTTCAAAAGAAATTTCAAAAGAAATTTCAAAAGAAATTTCAAAAG 1172
 QY 1103 ATGTTGGTGGAAACGAG-----ATGAGCAATATTTATGATGATGTTGTTGAT 1147
 Db 1173 ACATTGAGTGAAAGGGGATCGGTTGGGGAGAGACAGCGCTTGTCCAGATCTCAACAC 1232
 QY 1148 ATGTCCTTTAGCTTTGCTGTTTAAAGATCAATGGGTATGAAATTTTCCCGAGATCCAT 1207
 Db 1233 CACAGCCCTCGGCTGCGAATCTTTCGATGACGATACAAATTTTCTTCAGACGTTT 1292
 QY 1208 GGCTGAATTTACTAATGATTTAGCTTTGAAAGAGCAATATCGAGCTC----- 1254
 Db 1293 GAATAATTTCAAGATGAAGCGGCTTTCTTCTCTCTGCGGGCCAAACCCATGTCGA 1352
 QY 1255 -TTGAACAATATCATGCTCAAT 1311
 Db 1353 ATTGAGAAGCGTGTGTAATCTTTTCAAGAGCTTCCGACCTTGCATTTCTGACGAAAGAGC 1412
 QY 1312 AAATCTTGAAGTCACTGATTTCTTCAAGAGATTAATCACTGATTTCAACAGGCTTT 1371
 Db 1413 TATGAGCAGATCTAGAAATTTTCAAGAACCAATATCTTTAGAGGCGACTTGCACAGAAAT 1472
 QY 1372 CTAAATTAATCA-----CAAAGGCTGAAATCTCTTAAAGTTCCCTTATCAATPAC 1423
 Db 1473 CTCACCAATCAAACTATTTCAAGAGATTTGATGCTGTTGGAGTACCTTTGGACAT 1532
 QY 1424 CGGTTTAGAAGCGATATAACACTTAGACGAAATATACAGTTTATCAATGTAGACATACAG 1483

Db 1533 GAGTATCCACGCTTAGAAGCCAGAGTTATATGATTCATATGACGACAAATATGATG 1592
 QY 1484 AATCTGTAAGACTACATATCACTCATCAAAATATATATATATATATATATATATATATAT 1543
 Db 1593 GCAGAGGAGACTCTATATAGATGCCATCTTTGAGTAATCAAAATGTTTGAATGTC 1652
 QY 1544 TGTGAAAGATTTCACTGCTGCAATCTATTTATCGTGAAGATTTAAAGGCTTTGAAG 1603
 Db 1653 AAAATTGGACTTCAATATCGTACAATCTTTGATCAAGAGGAGTTGAAGCTTTCAACAAG 1712
 QY 1604 GTGGGTGTAGAGATAAGTTGACCGCTCAAGTTTGTAGGCAAAAGACCGCTTACTG 1663
 Db 1713 ATGGTGAAGGAATCCGGCATGGGATATATAATTTTCACTCCACACCGAGTGGGAGT 1772
 QY 1664 TTATTTCTGTTGCTGCAACACTTTTCTGCTCCGAAATTTATCAGATGCGGATTTTCAAT 1723
 Db 1773 TTATTTTTCATCAGCTACA-----TTTGAACCCGAATATTTCTGCCACTAGAATGCCCT 1826
 QY 1724 GGCACAAATGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1783
 Db 1827 CACAAAAATGTTGTTTCAAAATGCTTTTGAATGATGATGGCTGACATCTTTGCAACACT 1886
 QY 1784 CGATGAATGACCACTGATTCATGCTTTGAAATAATGAAATGATGATGATGATGATGATGAT 1843
 Db 1887 AGATGAATGAAAGTTTCACTGAGGAGTAAAGATGAGATGAGATGAGATGAGATGAGATGAG 1946
 QY 1844 TTGTTCTTCCAGACATGTTTCCGATTTTATTTTATGATTTAAAGATGCAATCTGTTGAT 1903
 Db 1947 ---GATTTCCAGATGATGCAAACTTTGCTTTAAAGTTTGGTTCAAATTAATGGAAGAAGT 2003
 QY 1904 TGAGATGAAGCTTTTAAATGCGACGCGGATGATTAAGTCCATGTTTATTTCAAACTG 1963
 Db 2004 AAATAATGATGTTGTTTAAAGGATACAGGAGCTGACATGCTCGCTCACATAAGAAAAACCTG 2063
 QY 1964 GTTGGAACTAACTAATGATGTTGAGAGAAGCTATATATGCAAGAGATGCTTTATGTC 2023
 Db 2064 GGAGTTGATCTCAATGTTTATGATCAAGAAAGGAGTGGCTTGAAGCCGGTATATACC 2123
 QY 2024 AACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2083
 Db 2124 AATTTTGAAGATGATCTTAAAGACTTATGCTATATCAGTAGGCTTTGGAACGCTGTACCT 2183
 QY 2084 GCGGCTATTTACTTTGTTGGGCCCCAAATTTATCAGAGGAGATTTGTTCAAGGCTCTGAATA 2143
 Db 2184 ACAACCAATACTACTAATGGGTGAGCTTGTGAAGATGATGTTGTTGAGAAAGTGCAC 2243
 QY 2144 TC---ATAATCTAATTAAGCTTAATGACACGCGAGGCTGACCTTCTTAAACGATATCCATAG 2200
 Db 2244 TCCCTCAATATGTTTGAAGCTTGTATCTTGAAGCTGCGACTGCGACTAACACACCAAAAC 2303
 QY 2201 CTTCAAGAGGGAATTTAAGGAAGGCAAAATTAACCGGTTAGCATGCTTTGATGATGATGATG 2260
 Db 2304 ATATCAGGCTGAAAGGCTCGAGGACAAAGGCTCAGGATAGCATGCTATATGAAGGA 2363
 QY 2261 AGAAGTGGAAAGTGAAGAGAGGTTGTGGAGGAT 2299
 Db 2364 TAATCCAGAGGCAACTGAGGAAGATGCCATTAAGACAT 2402

RESULT 12
 US-09-887-586A-43
 ; Sequence 43, Application US/09887586A
 ; Patent No. 6495354
 ; GENERAL INFORMATION:
 ; APPLICANT: Chappell, Joseph P.
 ; APPLICANT: No. 6495354, Joseph P.
 ; APPLICANT: Starks, Courtney M.
 ; APPLICANT: Manna, Kathleen R.
 ; TITLE OF INVENTION: SYNTHASES
 ; FILE REFERENCE: 07678-025001
 ; CURRENT APPLICATION NUMBER: US/09/887,586A
 ; FILING DATE: 2001-06-22
 ; PRIOR APPLICATION NUMBER: 09/398,395

; PRIOR FILING DATE: 1999-09-17
 ; PRIOR APPLICATION NUMBER: 60/130,628
 ; PRIOR FILING DATE: 1999-04-22
 ; PRIOR APPLICATION NUMBER: 60/150,262
 ; PRIOR FILING DATE: 1999-08-23
 ; NUMBER OF SEQ ID NOS: 58
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 43
 ; LENGTH: 2700
 ; TYPE: DNA
 ; ORGANISM: *Taxus brevifolia*
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (22)...(2607)
 ; OTHER INFORMATION: taxadiene synthase
 US-09-887-586A-43

Query Match 5.5%; Score 153.8; DB 4; Length 2700;
 Best Local Similarity 46.7%; Pred. No. 1.1e-29;
 Matches 943; Conservative 0; Mismatches 992; Indels 84; Gaps 11;

QY	335	AAATGTAGAAATTTCTGTTTCTTCATATGACACACATGGGTAGCCATGGT-----CCC 388
DB	414	AGACGGAGATTCAGTCCGTCGTGCATACGACACTCGTGGGTGGGAGGCTGGCGACCAT 473
QY	389	TTCTCCAAACTCACCCAAATCGCCTTGTTTCCCTCAGTGTCTCAATTTGGTTAAATTAATAA 448
DB	474	TTCTCTGATGATCTGAGAACCAACGGTTTCTCAGGCCCTCACTGGGTTTCAACA 533
QY	449	TCAGTTAATGATGTTTCATGGGGTCTGTTTAATCACACTCATATATCAATATCAATCAACCCGTT 508
DB	534	CCAGTCCAGGATGATCGTGGGGTATCGAATCGCACCTTTAGTTTATGGATCGATTGCT 593
QY	509	GCTTAAAGATCTCTATCTTCAACATTAGCATGTTATGTTGCAATTAAGATGAAATGT 568
DB	594	-----TACACGACCAATTTCTGTTATCGCCCTCTCGGTTTGGAAAC 635
QY	569	TGGGAAGATCAATAAATAAGGTCTAAGTTTATTTAGTCAATCTTGTCTCAGCTAC 628
DB	636	AGGCACACGCAAGTACAAAGGTGCTGAGTTTATTCAGAGATCTAAGNTACTCA 695
QY	629	TGAAAAAGTCAACATCTCCCAITGGTTTGGATCAATATTTCTGTTGCTTTGAGTA 688
DB	696	TEAGGAAGATGATGTTGTCCCC---GGATTTCAAAATAATCTTCTGCTCTGCTGCAAAA 752
QY	689	TGCAGAAAATTTGGACATAAACCTCTTTTCAAAACAAACAGATTTAGTTTGAATCTACA 748
DB	753	GGCAAAAGCTTTGGGATCAATCTTCCTTACGATCTTCCATTTATCAAAATATTTGTCGAC 812
QY	749	TAAGAGGAATTTGGAGCAAAAAGATGCCAT-----TCAATAGAGATGGATGATCTT 802
DB	813	AACACGGGAAGCCAGGCTTACAGATGTTTCTGGGCACGACACAATATTCAGCCAAAT 872
QY	803	GGCGTATATCTCTAGAGACTCGGTAATTTATATGATTTGGAATATGGTGAAGAAATATCA 862
DB	873	GTTGAATCGCTTGAAGGCTCTCGAGGAGTTATTGACTTGGAAACAAGATTATGAGGTTTCA 932
QY	863	GATCAAAAATGTTCTGTTTTCACATCCATCAGCAACAGCTGCTGCTTTCAATATCA 922
DB	933	AAGTAAAGATGATCTTTCTTCTGAGTCCCTCCCTCCACTGCTGTGATGATGAATATAC 992
QY	923	TCAAAATCTCGTTGCTCTTAATTTTAAATTTCACTTTTGGACAAGTTTGGTAAATGCAAT 982
DB	993	AGGGGACGAAAAATGTTTTCACCTTTCTCAACAATCTGCTGCACAAATTCGGCGGCTGGCT 1052
QY	983	CCCAACAGTTTATCCTCATGATTTTATTTATCCGACTTTCTATGTTTGAACAATTTGAAG 1042
DB	1053	GCCCTGTATGATTTCCATCGATCTGCTGGAACCGCTTTTCGCTGTTTGAATAACATTGACA 1112
QY	1043	ATTAGGAATTTTCAACCAATTTTCAAGTGGAAATTAAGAAATGTTTATGATGAACAATACAG 1102
DB	1113	TCTCGGAATCGGTCCCAATTTTCAACACAGAAATCAAGGAGCTCTTGTATTTATGTTCTACAG 1172

QY	1103	ATGTTGGGTGGAAACGAG-----ATGAGCAAAATATTTCATGATGTTGTAAAC 1147
DB	1173	ACATTTGGAGTGAAGGGGCATCGGTTGGGGCAGAGACAGCCTTGTTCAGATCTCAACAC 1232
QY	1148	ATGTGCTTTAGCCTTTTCGGTTATTAAAGGATCAATGGGTATGAAGTTTCCCGAGATCCATT 1207
DB	1233	CACAGCCCTCGGCTCGGAACCTTTTCGATGACCGATACATGTTTCTTCAGACGTTTT 1292
QY	1208	GGCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1254
DB	1293	GAATAATTTCAAGATGAACCGGGGTTCTTCCTCTCGGGGCCAAACCCATGTGCA 1352
QY	1255	-TTGAAACATATCATCGGTCACATATATATATACCAAGA--GGATTTATCTTCTGGAAC 1311
DB	1353	ATTGAGAACGTTGGTGAATCTTTTTCAGAGCTTCCGACCTTGCATTTCTCTGAGCAAGAGC 1412
QY	1312	AAATCTTGAAGTCAGCTGATTTCTTCAAGAGATAATATATCACTGATTCATAACAGGCTTT 1371
DB	1413	TATGACGATGCTAGAAATTTTCAGAACCATATCTTAGAGAGGCACTTGCACGAAAT 1472
QY	1372	CTAAATTAATTC-----CAAGAGGTGAAATGCTCTTAAGTTCCTTATCAATAC 1423
DB	1473	CTCAACCAATACAAACATATTTCAGAGATTTGAGTACGTTGGTGGATACCTTGGCAAT 1532
QY	1424	CGGTTTAGAACGCAATAACACTAGACGAAATATATACAGCTTTTACATGTAGCAATACAG 1483
DB	1533	GAGTATCCACGCTTAGAACGCAAGTTATATTGATTCATATGACGCAATATTATGTATG 1592
QY	1484	AAATCTGAAAACTCATATCACTCATCAAAATATTAGTAACACTGATTAACCTTAAGGTTGGC 1543
DB	1593	GCAGAGGAAGACTCTATATAGATGCCATCTTTGATTAATTCAAATGTTTAGAATTTGGC 1652
QY	1544	TGTTGAAGATTTTACACTCGCCTCAATCTATTATGTTGCAAGAAATTAAGGCTTTTGAAG 1603
DB	1653	AAAATTTGACTTCAATATCGTACATCTTTGATCAAGAGGATTTGAAGCTTCTAACAG 1712
QY	1604	GTGGTGTAGAGATAAGTTGGACCACTCAAGTTTCTAGGCAAAAGACCGCTACTG 1663
DB	1713	ATGGTGAAGGAATCCGGCATGGCAGATATAAATTTCACTCGACACCGAGTGGCGAGGT 1772
QY	1664	TTATTTCTCTGTTGTGCAACACTTTTCTCTCCGAAATTTATCAGATGGCGGTATTTCATG 1723
DB	1773	TTATTTTTCATCAGCTACA-----TTTGAACCCGAAATTTCTGCCACTAGAAATTTGCCCT 1826
QY	1724	GGCCAAAATGGCATATTAACATACAGTATGATGACCTTTTGTATGATTCGTTGGTGTACAT 1783
DB	1827	CACAAAATTTGGTTGTTTACAGTCTTTTTCAGATATGCTGACATCTTTTGCACACT 1886
QY	1784	CGATGAATTTGACCACTGATTTCAATGTTGAAAATGGAATGTAGATGTCGACAGGA 1843
DB	1887	AGATGAATTTGAAAATTTTCACTGAGGGAGTAAAGAGATGGGATACATCTTTTGTACATGA 1946
QY	1844	TTGTTGTTTTCAGAGATGTTTCGGATTTTATTTTATAGCATTTAAAGATGCAATCTGTTGGAT 1903
DB	1947	--GATTTCCAGAGTATGCAAACTTCTGTTTAAAGTTTGGTTTCAATTAATGGAAGAGT 2003
QY	1904	TGGAGATGAAGCTTTTAAATGGCAAGCGCGATGTAACATAGCCATGTTATTTCAAACTTG 1963
DB	2004	AAATAATGATGTTGTTTAAAGGTACAAAGGACGTGACATGCTGCTCACATAAGAAAACCTG 2063
QY	1964	GTTTGGAACTAATGATATGTTTGGAGAGCTATATGACAGAGATGCTTATGTC 2023
DB	2064	GGAGTTGATCTTCAATTTGTTATGATGACAGAAAGGAGTGGCTTGAAGCCGGTATATACC 2123
QY	2024	AACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2083
DB	2124	AACCTTTTGAAGAGTACTTAAGACTTATGCTATATCAGTAGGCTTGGACCGTGTACCT 2183
QY	2084	CGCGCTATTATCTTTGTTGGGGCCAAATTAATCAGAGAGATTTGTTGAAAGCTCTGAATA 2143
DB	2184	ACACCAATACTACTAATATGGTGGAGCTTGTGAAAGATGATGTTTCTTGAAGAAAGTGCATA 2243
QY	2144	TC---ATAATCTATTAACTAATGACGACGCGGGTCTGACTTCTTAAACGATATCCATAG 2200

Db 2244 TCCCTCAATATGTTGAGCTGTATCTTGGCTGGGACTTAACAAACGACACCAAAAC 2303
QY 2201 CTTCAAGAGGGAATTTAAGGAAGGCAATTAACCGGGTAGCAATTCATTTGAGTAACGG 2260
Db 2304 ATATCAGGCTGAAGAGGCTCGAGGACAAACAGCCTCAGGCATAGCATGTATATGAAGGA 2363
QY 2261 AGAAAGTGGAACTGGAGAGAGAGGTTCTGGAGGAGAT 2299
Db 2364 TTAATCCAGGAGCACTGAGGAAGATGCCATTAAGCAAT 2402

RESULT 13

US-09-895-752-43
; Sequence 43, Application US/09895752
; Patent No. 6559297
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 65592971, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/895,752
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/398,395
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/100,993
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 43
; LENGTH: 2700
; TYPE: DNA
; ORGANISM: Taxus brevifolia
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)....(2607)
; OTHER INFORMATION: taxadiene synthase
US-09-895-752-43

Query Match 5.5%; Score 153.8; DB 4; Length 2700;
Best Local Similarity 46.7%; Pred. No. 1.1e-29;
Matches 943; Conservative 0; Mismatches 992; Indels 84; Gaps 11;
QY 335 AATGTAGAAATTTCTTTTCATATGACAGCATGGGTAGCCATGGT-----CCC 388
Db 414 AGACGAGATATCAGTCCGCTCTGCATACGACATCTGGTGGGTGGCGAGCTGGCCACAT 473
QY 389 TTCTCCAAACTCACCAAAATCGCTTTGTTTCCCTGAGTGTCTCAATGGTTAATTAATAA 448
Db 474 TTCTCTGTGATGATCTGAGAAGCCACGGTTTCTCAGGCCCTCAACTGGGTTTTCACAA 533
QY 449 TCAGCTTAATGATGGTTCATGGGGTCTTGTTAATACACTCACTCAATAATCAATCAACCGTT 508
Db 534 CCAGCTCCAGGATGGATCGTGGGGTATCGAATCGCACTTTAGTTTATGCGATCGATTGCT 593
QY 509 GCTTAAAGATTTCTATCTTCAACATTTAGCATGTATTGTTGCAATTAATAAAGATGGAATGT 568
Db 594 -----THACAGCAACATTTCTGTTATCGCCTCTCGGTTTGGAAAC 635
QY 569 TGGGGAAGATCAATAAATAAAGTCTAAGTCTAAGTTTATTTAGTCAAAATCTTCCTCAGCTAC 628
Db 636 AGGGCAAGCAAGTACAAAGGCTGCTGAGTTTATTTGAGAGAAATCTAAGATTTACTCAA 695
QY 629 TGAAGAAGTCAACCACTCCCATTTGTTGACATCATATTTCTCGTTTGTGTTGAGTA 688
Db 696 TGAGGAAGATGATGTGTGCTCC-----GGATTTCCAAATTAATTTCTTCTGCTCTGCTGCAAA 752

QY 689 TCGGAAAAAATTGGACATAAAACCTCCTTTCAAAAACAAACAGATTTTAGTTGATGCTACA 748
Db 753 GSCAAAACGGTTGGGGATCAATCTTCTTACGATCTTCATTTTCAAAATATTTCGAC 812
QY 749 TAAAGAGGAATTTGGAGCAAAAAGATGCCAT-----TCAATGAGATGGATGATCTT 802
Db 813 AACACGGGAAGCAGGCTTTACAGATGTTTCTCGGCGAGCAGACAATTTTCAGCCAAAT 872
QY 803 GCGGTATATCTCTGAAGGACTCGGTAAATTTATATGATTGGAATATGGTGAAGAAATATCA 862
Db 873 GTTGAATGCGTTGGAAGGCTCTCGAGGAAGTTATTGACTGGAACAAGATTTATGAGTTTCA 932
QY 863 GATGAAAAATGCTTCTGTTTCAACTCAACATCAGACAGCTGCTGCTTTTCAATTAATCA 922
Db 933 AAGTAAAGATGATCTTCTCCTGAGCTCCCTCCCTCACTGCTGCTGTTACTGATGAATAC 992
QY 923 TCAAAATCTGTTGTTCTTAATTTAAATTTCACTTTTGGACAGTTTGGTAAATGCACT 982
Db 993 AGGGGACGAAAAATGTTTCACCTTTCTCAACAATCTGCTCGACAATTCGGCGGCTGCGT 1052
QY 983 CCCAACAGTTTATCTCATGATTTATTTATCCGACTTTCTATGCTTGGTGGACACAATTTGAAAG 1042
Db 1053 GCGCTGTATGATTCATCGATCTGCTGGAACGCTTTGCTGTTGATAAATCATTGAGCA 1112
QY 1043 ATTAGAATTTCAACCACTTTCAGAGTGGAAATTTAAATGTTTGTAGATGAACATACAG 1102
Db 1113 TCTCGAATCGGCTCGCCATTTCAACACAGAAATCAAGGAGCTCTTGTATGTTCTACAG 1172
QY 1103 ATGTTGGTGGAAACGAG-----ATGAGCAAAATTTTATCATGATGTTTGAAC 1147
Db 1173 ACATTGGAGTGAAGGGGCAATCGTTGGGCGAGACAGCCTTGTTCAGATCTCAACAC 1232
QY 1148 ATGTGCTTTAGCTTTTCGGTTATTTAGGATCAATGGTATGAAGTTTCCCCAGATCCATT 1207
Db 1233 CACAGCCTCGGCTCGAATCTTCCGATGACGAGTACAATGTTTCTTCAGACGTTTT 1292
QY 1208 GCGTGAATTTACTAATGAAATAGCTTTGAAAGAGAAATGAGCTC-----1254
Db 1293 GAATAATTTCAAGATGAAAACGGCGGTTCTTCTCTCGGGGCAAAACCAATGTCGA 1352
QY 1255 -TTGAAACATATCATCGCTCACATATATTATACCAAGA--GGATTTATCTTCTGGAAC 1311
Db 1353 ATTGAGAAGCGTGGTGAATCTTTTCAGAGCTTCGACCTTGCATTTCTTGACAAAGAGC 1412
QY 1312 AATCTTTGAAGTCAGCTGATTTCTCAAAGAGATAATATCCACTGATTTCAACAGAGCTTT 1371
Db 1413 TATGAGCAGTGTAGAAAATTTGCAGAACCAATCTTAGAGAGGCACTTTCGAACGAAAT 1472
QY 1372 CTAAATTAATTC-----CAAAGGTGGAAATGCTCTTAAGTTCCCTATCATATC 1423
Db 1473 CTCACCAATACAAAACCTATTCAAAGAGATTGAGTACGTGGTGGAGTACCCTTGGCACAT 1532
QY 1424 CGGTTTAGAACGATTAACACATAGACGAAATATACAGCTTTTACAATGTAGACAATAACAAG 1483
Db 1533 GAGTATCCCAAGCTTTAGAACCGAAGTTATATGATTCAATATGACGACAATTTATGTATG 1592
QY 1484 AATTTCTGAAAACCTACATATCATCTCAATCAATATTAGTAAACACTGATTTACCTAAGTTGGC 1543
Db 1593 GCAGAGGAAGACTCTATATAGATGCCATCTTTTGAATTAATTTCAAAATGTTTAGAATGGC 1652
QY 1544 TGTGTAAGATTTCTACCTCGCAATCTATTTATCTGTAAGAAATTAAGAGTTCTTGAAG 1603
Db 1653 AAAATTTGGAATCTCAATATCGTACAATCTTTGCAATCAAGAGGAGTTGAAGCTTTAAACAG 1712
QY 1604 GTGGGTGTAGAAATAAGTTGGACCAAGCTTCAAGTTTGTAGTCAAAAGACCGCTTACTG 1663
Db 1713 ATGTTGAAGGAATCCGGCATGCGAGATATAATTTCACTCGACACCGAGTGGCGAGGT 1772
QY 1664 TTAATTTCTGTTGTCGAACACTTTTCTGCTCCGGAATTTATGAGTGGCGGTTATTTCTG 1723
Db 1773 TTAATTTCTCATCAGCTACA-----TTTGAACCGGAATATTCTGCCCATGAAATTTGCTT 1826
QY 1724 GGCNAAAATGGCATTTAATTAACAGTAGTTGATGACTTTTTTTTGTATATCGTGGTACAAT 1783

```
Db 1827 CACAAAAATGGTGTGTTTCAAGTCTCTTTTATGATGATATGGCTGACATCTTTGGCAACT 1886
QY 1784 CGATGAATGACCAACTGATCAATCAATGCTTGAATAATGGAATGTAGATGTGCAACAGGA 1843
Db 1887 AGATGAATGAAAGTTTCACTGAGGAGTAAGAGATGGGATACATCTTTGCTACANGA 1946
QY 1844 TTGTTGTTTCAGAGCATGTTTCGGATTTTATTTTATAGCATTAATAAGATGCAATCTGTTGGAT 1903
Db 1947 --GATTCAGAGTGTATGCAAACTTGCTTTTAAAGTTTGGTTCAATTAATGGAAGAT 2003
QY 1904 TGGAGATGAAGCTTTTAAATGCGAGCGCGGATGTAACCTAGTGTATTCATAAAGTTG 1963
Db 2004 AAATAATGATGTGGTTAAGGTACAAGGACGTGACATGCTCGCTCACATAAGAAAAACCCCTG 2063
QY 1964 GTTGAACATAATGAATAGTATGTTGAGAGAACTATATGCAAGAGATGCTTATGTGCC 2023
Db 2064 GGAGTTGTACTTCAATGTTATGTACAGAAAGGAGTGCGTTGAAGCGGGTATATATACC 2123
QY 2024 AACATTAATGAATATATGAAAGCGCTTACGTGTCAATTCATAGGCCCGGATTTGCA 2083
Db 2124 AACTTTGAAGAGTACTTAAAGACTTATGCTATATACAGTAGGCCTTGGACCGTGTACCT 2183
QY 2084 GCCGCTATTACTTTGTTGGGCCCAATATATCAGAGAGATGTTGAAAGCTCTGAATA 2143
Db 2184 ACAACCAATACTACTAAATGCGGTGAGCTTGTGAAAGATGATGTTGTTGAGAAAGTGCACTA 2243
QY 2144 TC--ATAATCTATTAAAGCTAAATGAGCAGCGAGGCTGACTTCTAAACGATATCCATAG 2200
Db 2244 TCCCTCAATATGTTGAGCTTGATCTCTTGAGTGGCGACTTAAACACGACACCAAAAC 2303
QY 2201 CTTCAAGAGGAATTAAGGAAGGCAATTAACCGCGTAGCATTTGCAATTTGAGTAACGG 2260
Db 2304 ATATCAGGCTGAAAGGCTCGAGGACAAACAGCCTCAGGCATAGCATCTATATGAAGA 2363
QY 2261 AGAACTGGGAAGTGGAGAGAGGTTGTGGAGAGAT 2299
Db 2364 TAATCCAGGAGCAACTGAGGAAGATGCCATTAAGCACAT 2402
```

RESULT 14

```
US-09-903-012B-43
; Sequence 43, Application US/09903012B
; Patent No. 6569656
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 65696561, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/903,012B
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 60/100,993
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ IDS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 43
; LENGTH: 2700
; TYPE: DNA
; ORGANISM: Taxis brevifolia
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)...(2607)
; OTHER INFORMATION: taxadiene synthase
US-09-903-012B-43
```

Query Match 5.5%; Score 153.8; DB 4; Length 2700;
Best Local Similarity 46.7%; Pred. No. 1.1e-29;

```
Matches 943; Conservative 0; Mismatches 992; Indels 84; Gaps 11;
QY 335 AAATGTAGAAATTTCTGTTTTCATATGACACAGCATGGGTAGCCATGGT-----CC 388
Db 414 AGACGGAGATATACAGTCGGTCTGCATACGACACTGGTGGGTGGCGAGCTGGCGACCAT 473
QY 389 TTCTCCAAACTCACCAAAATCGCCTGTTTCCCTGAGTGTCTCAATTTGGTTAAATAA 448
Db 474 TTCTCTCATGATCTGGAAGCCACGGTTTCTCAGGCCCTCAATGGGTTTTCAACAA 533
QY 449 TCAGCTTAATGATGTTTCATGGGGTCTTGTAAATCACACTCATATCATATCAACCCGTT 508
Db 534 CCAGCTCCAGGATGATCGTGGGGTATCGAATCGCACTTTAGTTTATCGGATCGATTGCT 593
QY 509 GCTTAAAGATTTCTATCTTCAACATAGCATGATTTGTCATTTAAAGATGGAATGT 568
Db 594 -----TAAACAGCAAAATCTGTTATCGCCCTCTCGGTTTGGAAAC 635
QY 569 TGGGAAGATCAAAATAAAGTCTAAGTTTATTTAGTCAAAATCTTCTTCTAGCTAC 628
Db 636 AGGCGACAGCCAAAGTACAAAGGGTGTGAGTTTATTCAGAGAAATCTAAGATTACTCAA 695
QY 629 TCAAAAAAGTCAACCATCTCCCATTTGGTTTTCACATCATATTTCTGTTTCTTGAGTA 688
Db 696 TGAGGAAGATGATTTGTCCCC---GGATTTCCAAATAATCTTCTGCTGCTGCTGCAAAA 752
QY 689 TCGGAAAACTTGGACATAAACCCTCTTTTCAAAAAACAGATTTTATGTTTATGATCTACA 748
Db 753 GCGAAAAAGCGTTGGGGATCAATCTTCTTACGATCTTCCATTTATCAATATTTTCTCGAC 812
QY 749 TAAGAGGGAATTTGGAGCAAAAAAGATGCCAT-----TCAATGAGATGGATGATCTT 802
Db 813 AACACGGGAAGCCAGGCTTACAGATGTTTCTGGGCGACGACAAATATTCACGCCAACAT 872
QY 803 GGCCTATATCTCTGAAGACTCGGTAAATTTATATGATTGGAATATGGTGAAGAAATATCA 862
Db 873 GTTGAATGCGTTGGAAGGTCTCGAGGAGTTATTTGACTGGAAACAGATTTATGAGTTTCA 932
QY 863 GATGAAAAATGGTCTGTTTTCACCTCAACATCAGCAACAGCTGCTGCTTTTCATTAATCA 922
Db 933 AAGTAAAGATGGATCTTTCTCTGAGTCCCTCCCTCCACTGCTGCTGTGTACTGATGAATAC 992
QY 923 TCAAAATCCTGTTGTTCTTAATTTTAAATTTCACTTTTGGACAGTTTGGTAATGAGT 982
Db 993 AGGGACGAAAAATGTTTCACTTTTCAACAAATCTGCTGACAAATTCGGCGGCTCGGT 1052
QY 983 CCCAACAGTTTATCTCATGATTTATTTATCCGACTTTTCTATGTTTGCACAAATTTGAAAG 1042
Db 1053 GCCCTGTATGTTATTCATCGATCTGCTGGAAGCGCTTTGCTGTTGATACATTGAGCA 1112
QY 1043 ATTAGGAATTTACACCATTTTCAGAGTGGAAATTAATAATGTTTATGATGAAACATACAG 1102
Db 1113 TCTCGGAATCGGTGCGCAATTTCAAAACAGAAATCAAAGAGAGCTCTTGATTATGCTACAG 1172
QY 1103 ATGTTGGGTGGAACGAG-----ATGACCAATATTTATGATGATGTTGTTAAC 1147
Db 1173 ACATTGAGGTGAAGAGGGGATCGGTTGGGCGACAGACAGCCTTGTTCAGACTCTCAACAC 1232
QY 1148 ATGTGCTTTAGCCCTTTTCGGTTTATTAAGGATCAATGGGTATGAAAGTTTCCCAGATCCATT 1207
Db 1233 CACAGCCCTCGGCTCGAACTCTTCGATGCGACGGATACAATGTTTCTTCAGACGTTTTT 1292
QY 1208 GGCTGAATTAAGTAATAGCTTTTGAAGACGTAATATGCAAGTCT-----1254
Db 1293 GAATAAATTTCAAAGATGAATAACGGGGGTTCTTCTCTCTCGGGGCCAAACCCCATGTCGA 1352
QY 1255 -TTGAAACATATCATCGTGCATATATATATACAGA--GGATTATCTTCTGGAAGAAC 1311
Db 1353 ATTGAGAAAGCGTGGTGAATCTTTTCAGAGCTTCGACCTTTCGATTTCTGACGAAAGAGC 1412
QY 1312 AAATCTTTGAGTCAAGTCTGATTTCTCTCAAGAGATATATATCCACTGATTTCAACAGGGTTT 1371
Db 1413 TATGGACGATGCTAGAAAATTTGAGAACCATATCTTAGAGGACACTTCCACAGAAAT 1472
```

QY	1372	CTAATTAATTCA-----CAAGAGGTGAAATGCTCTTAAGTTCCCTTATCAATAC	1423
Db	1473	CTCAACCAATACAAACTATTCAAGAGATGAGTACGTTGGGAGTACCTTTGGCAGAT	1532
QY	1424	CGGTTTAGAACGCAATAACACATAGACGAAATATACAGCTTTTCAATGTAGACAAATACAG	1483
Db	1533	GAGTATCCACCGCTTAGAGCCGAGATTAATGATTCATATGACGACAAATATGATG	1592
QY	1484	AATTCGAAATACATCACTCAATCAATATTAATAGTAACTGATTAAGTTGGC	1543
Db	1593	CGACGAGGAGACTCTATATAGAAATGCTTTGAGTAATTCAAATATGTTAGAAATGGC	1652
QY	1544	TGTTGAGATTTCTACACTGCGCAATCTATTTATGTTGAGATTAATAAGGCTTTGAAAG	1603
Db	1653	AAATTTGAGACTCAATATCTGTAATCTTTGATCAAGAGGAGTTGAGCTTTCAACAG	1712
QY	1604	GTGGGTGGTAGAATAAGTTGGACCGAGCTCAAGTTTGTAGGCAAAAGACCGCTTACTG	1663
Db	1713	ATGGTGAAGGATCCGGCATGGCAGATATAATTTCACTCGACACCGAGTGGCGAGGT	1772
QY	1664	TTATTTCTGTTGTCGCAACATTTCTGCTCCGAAATATCAAGATGCGCGTATTTCAATG	1723
Db	1773	TTATTTTTCATCAGCTACA-----TTTGAACCCGAAATATCTGCCACTAGAAATGSCCT	1826
QY	1724	GGCCAAATGCGATATTAACACTAGTGTGATGACTTTTGTATCGTGGTACAAT	1783
Db	1827	CACAAATATGTTGTTTCAAGTCTTTTGAATGATGCTGATGATCTTTGCAACACT	1886
QY	1784	CGATGAATTCACCAACCTGATTCATATGTTGTAATAATGAAATGATGATGTCGACAAAGA	1843
Db	1887	AGATGAATTAAGAGTTTCACTGAGGAGTAAAGAGATGGATACATCTTTGTCATATGA	1946
QY	1844	TTGTTGTCAGAGCATGTTGCGATTTTATTTTAGCAATTAATAAGATGCAATCTGTTGGAT	1903
Db	1947	---GATTCAGAGTATGCAAACTTCTTTAAAGTTTGGTTCAATTAATGGAAGAAGT	2003
QY	1904	TGGAGATGAGCTTTTAAATGGCAAGCGCGAGTAACTAGCCATGTTATTCAAACTTG	1963
Db	2004	AAATATGATGTTGGTTAAGGTACAGGACGTGACATGCTGCTCAATAGAAACCCCTG	2063
QY	1964	GTTGGAACAAATGAATGATGTTGAGAGAGCTATATGACAAAGAGATGCTTTATGTCC	2023
Db	2064	GGAATGTTACTTCAATGTTTATGACAAAGAGGAGTGGCTTGAAGCGGGTATATACC	2123
QY	2024	AACTTAATGAATATGAAACGCTTACGTGTCATTTGCTATGATGAGCCCGATTTGCAA	2083
Db	2124	AACTTTTGAAGTACTTAAAGACTTATGCTATATCAGTAGGCTTTGGAACCGTGTACCCCT	2183
QY	2084	CGCGCTATTTACTTTGTTGGGGCCCAATTTATCAGAGGAGATGTTGMAAGCTCTGAATA	2143
Db	2184	ACAACAAATACTAATGAGGTGAGCTTTGAAAGATGATGTTGTTGAGAAAGTGACTA	2243
QY	2144	TC---ATAATCTATTAAGCTAATGACGACGCGGGTTCGCTTTCTTAAACGATATCCATAG	2200
Db	2244	TCCCTCAATATGTTGAGCTTGTATCTTTGAGCTGGCGACTAACCAACGACACCAAAAC	2303
QY	2201	CTTCAAGAGGAATTTAAGAAAGGCAATTAACGCGGTAGCATTCATTTGAGTACGG	2260
Db	2304	ATATCAGGCTGAAAAGGCTCGAGGACCAACAAAGCCTCAGGCAATGATGATATGAAGGA	2363
QY	2261	AGAAAGTGGAAATGGAAGAGGTTGTTGGAGGAGAT	2299
Db	2364	TAATCCAGGAGCACTGAGGAGATGCCATTAAGCAT	2402

RESULT 15
US-09-593-253-1
; Sequence 1, Application US/09593253
; Patent No. 6610527
; GENERAL INFORMATION:
; APPLICANT: Rodney B. Croteau, Mark R. Wildung
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TAXOL

BIOSYNTHESIS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Donald L. Stephens Jr.
STREET: One World Trade Center
121 S.W. Salmon Street
Suite 1600
CITY: Portland
STATE: Oregon
COUNTRY: United States of America
ZIP: 97204
COMPUTER READABLE FORM:
MEDIUM TYPE: Disk, 3-1/2 inch
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/593,253
FILING DATE: 13-Jun-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: 08/843,363
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Donald L. Stephens Jr.
REGISTRATION NUMBER: 34,022
REFERENCE/DOCKET NUMBER: 4630-46842/DLS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (503) 226-7391
TELEFAX: (503) 228-9446
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2700 base pairs
TYPE: nucleic acid
STRANDEDNESS: double stranded
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-593-253-1

atch	5.5%;	Score 153.8;	DB 4;	Length 2700;
cal Similarity	46.7%;	Pred. No. 1.1e-29;		
943;	Conservative	0;	Mismatches 992;	Indels 84; Gaps 11;
335	AAATGTAGAAATTTCTGTTTCTT	CATATGACAGCATGGTAGCCATGGT	-----CCC	388
414	AGACGGAGATATCAGTCCGCTCT	GCATACGACACTCGTGGGTGGCGAGGCTGGCGACCAT	473	
389	TTCTCCAAACTACCCAAATCGCC	TGTTTCCCTGAGTGTCTCAATTTGGTTAATTAATAA	448	
474	TTCTCTGATGGATCTGAGAAGCC	ACGGTTTCTTCAGGCGCTCACTGGGTTTTCACAA	533	
449	TCAGCTTAATGATGGTTCATGGG	GTCTGTTAAATCACACTCATAATCAATAATCAACCCGTT	508	
534	CCAGCTCCAGATGAGATCGTGGG	TATCGAATCGCACTTTAGTTTATGCGATCGATTGCT	593	
509	GCTTAAAGATTTCTCTATCTTCA	ACATTAGCATGATTTGTCATTAATAAGATGGAATGT	568	
594	-----TAA	CACGCAATTCGTTATCGCCCTCTCGGTTTGGAAAA	635	
569	TGGGAAGATCAATAATAAGGTCT	AAGTTTATGATCAAACTTTCGTTTCAGCTAC	628	
636	AGGGCAGCAGCATCAACAGGTG	CTGAGTTTATTTGCAGAGAACTTAAGATTACTCAA	695	
629	TGAAAAAAGTCAACCATCTCCAT	TGGTTTGGACATCATATTTCCCTGGTTTCTGCTGAGTA	688	
696	TGAGGAAGATGAGTTGTTCCCTC	---GGATTTCCAAATTAATCTTTCCTGCTCTGCTGCAAAA	752	
689	TGCGAAAAAATTTGGACATAAAC	CTCCTTTCAAAACAAACAGATTTTAGTTTGAATGCTACA	748	
753	GGCAAAAGCGTTGGGATCAATCT	TCTTACGATCTTCCATTTATCAAAATATTTGTCGAC	812	
749	TAAGAGGGAATTTGGACAAAAA	AGATGCCAT-----TCAAATGAGATGGATGATCTT	802	

Db	813	AAACGGGAAGCCAGGCTTTACAGATGTTTCTGCGCGACGACACAATAATTTCCAGCGAACAT	872
Qy	803	GGCGTATATCTCTGAAGGACTCGGTAAATTTATATGATTGGAATATGGTGAAGAAATATCA	862
Db	873	GTGGAATCGTTGGGAAGTCTCGAGGAAGTTATTGACTGGAAACAGAAATATGAGTTTCA	932
Qy	863	GATGAAAAATGGTTCGTGTTTCAACTCAACATCAGCAACAGCTGCTGTTTCATTAAATCA	922
Db	933	AAGTAAAGATGGATCTTTCCCTGAGCTCCCTCGCTCCACTGCCTGTGTACTGATGAATAC	992
Qy	923	TCAAAATCCTGTTGTTCTTAATTAATTAATTTCACTTTTGGACAAGTTTGGTAAATGCAGT	982
Db	993	AGGGGACGAAAAATGTTTCACTTTTCTCAACAACTGCTGCACAAATTCGCGCGCTCGT	1052
Qy	983	CCCAACAGTTTTATCCTCATGATTTATTTATCCGACTTCTATGTTGTCACACAATTTGAAG	1042
Db	1053	GCCCTGTATGTAATTCATCGAATCTGCTGGAAACGCTTTGCGTGGTGAATACATTGAGCA	1112
Qy	1043	ATTAGGAATTTACACACATTTTCAGAGTGGAAATTAATAATGTTTATAGTGAACAATACAG	1102
Db	1113	TCTCGGAATCGGTGCGCAATTTCAAACAAGAAATCAAAGGAGCTCTTGATTATGTTACAG	1172
Qy	1103	ATGTTGGGTGAAGCGAG-----ATGAGCAAAATATTCATGATGTTGTTAAC	1147
Db	1173	ACATGGAGTGAAGGGGCAATCGGTTGGGCGAGAGACAGCGCTTGTTCAGATCTCAACAC	1232
Qy	1148	ATGTGCTTTAGCCTTTCCGTTTATTAAGGATCAATGGGTATCAAGTTTCCCGACATCCATT	1207
Db	1233	CACAGCCTCGGCTCGGAATCTCTTCGATCGCAGGATACAAATGTTTCTTCAGACGTTTT	1292
Qy	1208	GGCTGAATTTACTAAATGAATTAGCTTTGAAAGAGCAATATGCGAGCTC-----	1254
Db	1293	GAATTAATTTCAAAGATGAAACGCGCGGTTCTTCTCCTCTGCGGGCCAAACCCATGTCGA	1352
Qy	1255	TTCAAAACATATCATGCTGCTACATATATATACCAAGA--GGATTTATCTTTCGGAAGAC	1311
Db	1353	ATTGGAAGCGTGTGTAATCTTTTCAGAGCTTCGACCTTGCATTTCTGACGAAAGAGC	1412
Qy	1312	AAATCTTGAAGTCAGCTGATTTCTCAAAGAGATATATCCACTGATTTCAACACAGGCTTT	1371
Db	1413	TATGGACGATGCTAGAAAAATTTGCAGAAACCATATCTTAGAGAGGCATTCGACGAAAT	1472
Qy	1372	CTAAATTTAATTCA-----CAAAGAGTGGAAAAATGCTTCTTAAGTTCCCTATCAATAC	1423
Db	1473	CTCAACCAATACAAAACTATTCAAAGAGATGAGTAGCTGCTGGTGGAGTACCTTTGGCACAT	1532
Qy	1424	CGGTTTAGAAGCGATAAACACTAGCGAAATATACAGCTTTTCAATGTAGACAAATACAG	1483
Db	1533	GAGTATCCCAAGCTTTAGAACGGAAGTTATATGATTTCAATATGACGACAAATATGATG	1592
Qy	1484	AATTTCTGAAAACATCATATCACTCATCAAAATATAGTAAACATGATTTACTGAAGTTGCG	1543
Db	1593	GCAGAGGAAGACTCTATATAGAATGCCATCTTTGAGTAAATCAAATGTTTAGAATGGC	1652
Qy	1544	TGTTGAAGATTTCTACACCTGCCAATCTATTTATCGTGAAGAAATTAAGAGTCTTGAAG	1603
Db	1653	AAAATTTGGACTTCAATATCGTACAATCTTTTGCATCAAGAGGAGTTGAAGCTTCAACAAG	1712
Qy	1604	GTGGTGGTAGAGAAATAAGTTTGACACGAGCTCAAGTTTGTCTAGGCAAAAGACCGCTACTG	1663
Db	1713	ATGGTGAAGGAATCCGGCATGGCAGATATAAATTTCACTCGACCCGAGTGGCGAGGT	1772
Qy	1664	TTATTTCTCTGTGCTGCAACACTTTTCGCTCCCGAAATTTATCAGATCGCGTATTTCAATG	1723
Db	1773	TTATTTTTCATCAGCTACA-----TTTGAAACCGGAATATCTGCGCATGAGAAATTCGCTTT	1826
Qy	1724	GGCCAAAAATGGCAATTTAACTACATGATGTAAGTATTTTTTGTATATCCGTTGGTACAAT	1783
Db	1827	CACAAAAATTTGGTTGTTTACAGTCTCTTTTGATGATATGGCTGACATCTTTTGCAACACT	1886
Qy	1784	CGATGAATTAGCCAACTGATTTCAAATGTTTGAATAATGAATGATGATGTCGACAGAAG	1843
Db	1887	AGATGAATTTGAAAGTTTCACTGAGGAGTAAAGAGATGGGATACATCTTTGCTACTACGA	1946

QY	1844	TTGTGTGTT	CAGAGCATGTT	CGAGTATTTT	TAATTTT	TAGCATTA	AAAGATG	CAATCTG	TGCGAT	1903				
Db	1947	---	GATTC	CCAGG	TGTATG	CAAACTG	CTTTT	AAAAGT	TGGTT	CAAAATTA	TGGAAGAAGT	2003		
QY	1904	TGGAGATG	GAAGCTTTT	AAATGG	CAAGCGCG	GATGTA	ACTAG	CCATGTTAT	TAAACTTG	1963				
Db	2004	AAATAT	ATGATG	TGGTT	AGGTG	TAAGGAC	GTGCAT	GCTGC	TACATAT	AGAAAAC	CCCTG	2063		
QY	1964	GTTGGAA	CTAAT	GAATAG	TATGTT	TGAGAG	AAGCTAT	TATGG	ACAAGAGATG	CTTATG	TGTGCC	2023		
Db	2064	GGAGTTG	TACTCA	ATTGTT	TATG	TACAG	AAAGGAGT	TGGCTT	GAA	CCGGGTAT	TATACC	2123		
QY	2024	AACATTA	AAATGA	ATATAT	TGNA	AACGCTT	AGCTGT	CAATTTG	CAATAG	CGCGGAT	TGTCAA	2083		
Db	2124	AAC	TTTTGA	GAGTACTT	TAAG	ACTTAT	GTCTAT	TATC	ATAGT	AGGCTT	TGGACG	GTGTAC	2183	
QY	2084	GCCGGCT	ATTATTT	TACTTT	TGTGG	GGGCCAA	ATATAT	CAGAG	GAGATG	TGTC	CAAAAGCT	CTGTA	2143	
Db	2184	ACAACCA	ATACT	ATAAT	TGGT	GAGCTTG	TGCA	AAAGATG	ATGTTG	TTC	TGAG	AAAGTG	CACTA	2243
QY	2144	TC---	ATAAT	CTATTT	AA	GCTAT	TGAC	ACG	GGTFC	GTCTT	CTAA	CGAT	TCC	2300
Db	2244	TCCCT	CAAAAT	TGTTT	GAGCTT	GTAT	CTCTT	GAGCT	TGGCG	ACTAAC	CAAC	GAC	CA	2303
QY	2201	CTTCA	GAGGGA	ATTTT	AA	GGAAG	CGCAAT	TAA	ACCGG	GTAGCT	TGCATTT	TGAGT	AT	2360
Db	2304	ATA	T	CAGGCT	GAAA	AGSCT	CGAG	GAC	CA	CAAG	CCCTC	AGG	CA	2363
QY	2261	AGAA	GTGG	GAAGT	TGGA	GAAG	AGGTT	TG	GAGG	AGAT	2299			
Db	2364	TAAT	CCAG	GAGCA	ACTG	AGGA	AGATG	CCAT	TA	AG	CA	2402		

Search completed: October 17, 2004, 03:30:43
Job time : 238.559 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 16:35:39 ; Search time 1304.77 Seconds
(without alignments)
11232.937 Million cell updates/sec

Title: US-10-041-018-361
Perfect score: 2792
Sequence: 1 cccactcctttatcac.....aaaaaaaaaaaaaaaaaaaaa 2792

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_238sep04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001as.*
5: Geneseq2001bs.*
6: Geneseq2002as.*
7: Geneseq2002bs.*
8: Geneseq2003as.*
9: Geneseq2003bs.*
10: Geneseq2003cs.*
11: Geneseq2003ds.*
12: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2792	100.0	2792	12	Adm98941 Diterpene
2	2582.6	92.5	3117	12	Adm98942 Diterpene
3	1428.6	51.2	2594	12	Adm98948 Diterpene
4	809.4	29.0	2638	12	Adm98947 Diterpene
5	782.2	28.0	2658	12	Adm98958 Diterpene
6	782.2	28.0	2658	12	Adm98958 Diterpene
7	780.6	28.0	2658	2	Adm98961 Diterpene
8	732.4	26.2	2506	12	Aat60339 Pumpkin e
9	453.2	16.2	2403	12	Adm98957 Diterpene
10	416.8	14.9	2154	8	Adm94223 Corn Ent-
11	379.4	13.6	2223	8	Ada70130 Rice gene
12	379.4	13.6	2223	12	Ada69971 Rice gene
13	330	11.8	1559	12	Adi45176 Rice isop
14	329.4	11.8	1875	8	Adm98955 Diterpene
15	315.6	11.3	2193	8	Ada70068 Rice gene
16	241.6	8.7	2313	8	Ada69972 Rice gene
17	238.2	8.5	1157	12	Ada70298 Rice gene
18	194.8	7.0	2861	3	Adm94233 Corn Ent-
19	194.8	7.0	2861	12	Aaa38937 Grand fir
20	172.8	6.2	439	12	Adm98945 Diterpene
21	172	6.2	577	12	Adi45560 Wheat iso

22	153.8	5.5	2700	2	AAT97447 Pacific Y
23	153.8	5.5	2700	3	Aaa38931 Yew taxad
24	153.8	5.5	2700	12	Adm98943 Diterpene
25	139.2	5.0	531	12	Adm94227 Wheat Ent
26	133.8	4.8	2705	12	Adm98977 Diterpene
27	130.6	4.7	468	12	Adm94225 Rice Ent-
28	130	4.7	339	12	Adm94225 Rice Ent-
29	128.4	4.6	314	12	Adm94225 Rice Ent-
30	120.8	4.3	2528	2	Adm94225 Rice Ent-
31	119.2	4.3	2424	2	Aax87531 B-alpha-b
32	119.2	4.3	2424	3	Aaa38932 Grand fir
33	119.2	4.3	2424	4	Aaf73382 Grand fir
34	119.2	4.3	2525	2	Aax87530 E-alpha-b
35	119.2	4.3	2525	2	Aax87529 Grand fir
36	119.2	4.3	2571	2	Aax87529 Grand fir
37	119.2	4.3	80374	12	Aax87504 Grand fir
38	119	4.3	632	5	Adm98956 Diterpene
39	118.4	4.2	426	12	Aah87659 Peppermin
40	116.8	4.2	463	12	Adm98949 Diterpene
41	113.4	4.1	2460	3	Aaz99415 Spybean c
42	113.4	4.1	2460	6	Aad40203 Soybean c
43	112.4	4.0	441	12	Adm94221 Soybean E
44	100.6	3.6	2543	12	Adm94221 Soybean E
45	99.8	3.6	230	12	Adm94221 Soybean E

ALIGNMENTS

RESULT 1
ADM98941
ID ADM98941 standard; DNA; 2792 BP.
XX
AC ADM98941;
XX
DT 01-JUL-2004 (first entry)
XX
DE Diterpene synthase DNA #1.
XX
KW Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;
KW diterpene precursor; diterpene synthase; defence toxin;
KW volatile defensive signal; pollinator attractant; photoprotectant.
XX
OS Stevia rebaudiana.
XX
FN US2004072323-A1.
XX
PD 15-APR-2004.
XX
PF 07-JAN-2002; 2002US-00041018.
XX
PR 05-JAN-2001; 2001US-0259880P.
XX
PA (MATS/) MATSUDA S P T.
PA (HART/) HART E A.
XX
PI Matsuda SPT, Hart EA;
XX
DR WPI; 2004-373921/35.
XX
PT New unicellular organisms comprising exogenous nucleic acids encoding a
PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for
PT producing diterpenes and diterpene precursors.
XX
PS Claim 3; SEQ ID NO 361; 38pp; English.
XX
CC The invention relates to a unicellular organism for producing a diterpene
CC or diterpene precursor comprising an exogenous nucleic acid sequence
CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene

CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the
CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC microorganism encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents DNA encoding a diterpene synthase polypeptide of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification but was obtained in electronic format from
CC USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 2792 BP; 911 A; 474 C; 550 G; 857 T; 0 U; 0 Other;

Query Match 100.0%; Score 2792; DB 12; Length 2792;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2792; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCCACTCATCCCTTATCAACCAATACCATCGTTCTGCCACCGAGAGACTGATACGCGAG 60
DB 1 CCCACTCATCCCTTATCAACCAATACCATCGTTCTGCCACCGAGAGACTGATACGCGAG 60
QY 61 GACCTGACGACGCTTTAATCTCTGTTGGACCAACCTTAAACAAATTCGACCCATCATCT 120
DB 61 GACCTGACGACGCTTTAATCTCTGTTGGACCAACCTTAAACAAATTCGACCCATCATCT 120
QY 121 ACTCCGGTGACGAGTTTCAATTTAGTAAATCAGCAATTAACCAATCAATCAATCAATC 180
DB 121 ACTCCGGTGACGAGTTTCAATTTAGTAAATCAGCAATTAACCAATCAATCAATCAATC 180
QY 181 TTTCACTATGATCGCGTCCCTTTCTGTTAAACCAATCAATCAATCAATCAATCAATCAAT 240
DB 181 TTTCACTATGATCGCGTCCCTTTCTGTTAAACCAATCAATCAATCAATCAATCAATCAAT 240
QY 241 CTATTCATACAGCATCAACTTCAATGTTGACCAAACTAATCCCACTAATCTCATTTG 300
DB 241 CTATTCATACAGCATCAACTTCAATGTTGACCAAACTAATCCCACTAATCTCATTTG 300
QY 301 ATACACCAAGACGATCCAAACAGTTTAAATGTAAGATTTAGTAAATTTCTGTTTCTCAT 360
DB 301 ATACACCAAGACGATCCAAACAGTTTAAATGTAAGATTTAGTAAATTTCTGTTTCTCAT 360
QY 361 ATGACACAGATCGGTAGCATGCTCCCTTTCTCCAAACTCACCCAAATCGCTTGTTC 420
DB 361 ATGACACAGATCGGTAGCATGCTCCCTTTCTCCAAACTCACCCAAATCGCTTGTTC 420
QY 421 CTGAGTGTCTCAATTTGTTAATTAATCAGCTTAATGATGTTTCATGCGGTCTTGT 480
DB 421 CTGAGTGTCTCAATTTGTTAATTAATCAGCTTAATGATGTTTCATGCGGTCTTGT 480
QY 481 ATCACACTATATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 540
DB 481 ATCACACTATATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 540
QY 541 GTATTGTTGATTAATAAAGATGGAATGTTGGGGAAGATCAATAATAAAGGTCTAAGTT 600
DB 541 GTATTGTTGATTAATAAAGATGGAATGTTGGGGAAGATCAATAATAAAGGTCTAAGTT 600
QY 601 TTATTGATCAATCTTGTCTGATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 660
DB 601 TTATTGATCAATCTTGTCTGATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 660
QY 661 ACATCATATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 720
DB 661 ACATCATATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 720
QY 721 AACCAACAGATTTTAGTTTGTGCTTACATAGAGGGAATTTGGAGCAAAAAGATGCCATT 780

DB AACCAACAGATTTTAGTTTGTGCTTACATAGAGGGAATTTGGAGCAAAAAGATGCCATT 780
QY CAATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
DB CAATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
QY GGAATATGTTGGAAGAAATATCAGATGAAATGTTCTGTTTCACTCAACCTACCAACAG 900
DB GGAATATGTTGGAAGAAATATCAGATGAAATGTTCTGTTTCACTCAACCTACCAACAG 900
QY CAGCTGCTGCTTCAATTAATCATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 960
DB CAGCTGCTGCTTCAATTAATCATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 960
QY TGGCAAGATTTGGTAAATGAGTCCCAACAGATTTATCCTCATGATTTATTTATCCGACT 1020
DB TGGCAAGATTTGGTAAATGAGTCCCAACAGATTTATCCTCATGATTTATTTATCCGACT 1020
QY CTATGTTGACAAATTTGAAGATTTAGAAATTTAGAAATTTAGAAATTTAGAAATTTAGAA 1080
DB CTATGTTGACAAATTTGAAGATTTAGAAATTTAGAAATTTAGAAATTTAGAAATTTAGAA 1080
QY ATGTTTTAGATGAAATACATGATGTTGGTGGACGAGATGAGCAAAATATTTATCATGAT 1140
DB ATGTTTTAGATGAAATACATGATGTTGGTGGACGAGATGAGCAAAATATTTATCATGAT 1140
QY TTTGTAACATGCTTTAGCTTTTGGTTTAAAGATCAATGGTATGAAGTTTCCCGAG 1200
DB TTTGTAACATGCTTTAGCTTTTGGTTTAAAGATCAATGGTATGAAGTTTCCCGAG 1200
QY ATCCATTTGGCTGAAATTTAATGATTTAGCTTTGAAAGACGAAATATGAGCTTTTGA 1260
DB ATCCATTTGGCTGAAATTTAATGATTTAGCTTTGAAAGACGAAATATGAGCTTTTGA 1260
QY CATATCATGCTGACATATATATATCAAGAGATTTATCTTCTGCAAAACAAATCTTGA 1320
DB CATATCATGCTGACATATATATATCAAGAGATTTATCTTCTGCAAAACAAATCTTGA 1320
QY AGTCAGCTGATTTCTTCAAGAGATTAATATCACTGATTTCAACAGGCTTTCTTAAATTA 1380
DB AGTCAGCTGATTTCTTCAAGAGATTAATATCACTGATTTCAACAGGCTTTCTTAAATTA 1380
QY TTCAACAAAGAGTGGAAATGCTTTAAGTTCCCTATCAATCCGTTTGAAGCGGATA 1440
DB TTCAACAAAGAGTGGAAATGCTTTAAGTTCCCTATCAATCCGTTTGAAGCGGATA 1440
QY AACTAGACGAAATATACAGCTTTTCAATGTAGCAATTAACAGAAATTTCTGAAATACAT 1500
DB AACTAGACGAAATATACAGCTTTTCAATGTAGCAATTAACAGAAATTTCTGAAATACAT 1500
QY ATCACTCATCAAAATTTAGTAACTGATTTAAGTTGCTGTTGAAAGATTTCTTACA 1560
DB ATCACTCATCAAAATTTAGTAACTGATTTAAGTTGCTGTTGAAAGATTTCTTACA 1560
QY CTTGGCAATCTATTTATCGTGAAGATTAAGGTTCTGAAAGTGGTGGTAGAATA 1620
DB CTTGGCAATCTATTTATCGTGAAGATTAAGGTTCTGAAAGTGGTGGTAGAATA 1620
QY AGTTGGACCAAGCTCAAGTTTGTAGGCAAAAGACCGCTACTGTTATTTCTGTTGCTG 1680
DB AGTTGGACCAAGCTCAAGTTTGTAGGCAAAAGACCGCTACTGTTATTTCTGTTGCTG 1680
QY CAACACTTTTCTGTTCCGAAATTTATCAGATCGGCTATTTTATGGGCCAAAAGATG 1740
DB CAACACTTTTCTGTTCCGAAATTTATCAGATCGGCTATTTTATGGGCCAAAAGATG 1740
QY TAACACTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800
DB TAACACTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800
QY TGATTTCAATGTTGAAATTTAGATGTTAGATGTTGCAAGGATTTGTTGTTGAGAGATG 1860
DB TGATTTCAATGTTGAAATTTAGATGTTAGATGTTGCAAGGATTTGTTGTTGAGAGATG 1860

Db 1801 TGATTCATGTTGTGAAATGGAATGTAGATGTGCAAGGATTTGTTCTCAGAGCATG 1860
QY 1861 TTCGGATTTTATTTTATAGCATTAAAGATGCAATCTGTTGGATTGAGATGAAGCTTTTA 1920
Db 1861 TTCGGATTTTATTTTATAGCATTAAAGATGCAATCTGTTGGATTGAGATGAAGCTTTTA 1920
QY 1921 AATGGCAAGCGCGGATGATAGCATTGTTATTCAAACTTGGTTGGAACATAAAGAATA 1980
Db 1921 AATGGCAAGCGCGGATGATAGCATTGTTATTCAAACTTGGTTGGAACATAAAGAATA 1980
QY 1981 GTATGTTGAGAGAGATATATGCAAGAGATGCTTATGTCACATTAATAAGATATA 2040
Db 1981 GTATGTTGAGAGAGATATATGCAAGAGATGCTTATGTCACATTAATAAGATATA 2040
QY 2041 TCGAAAAACGCTTACGTTGATTTGATTTAGGCGCGGATTTCAAGCGGCTATTTACTTTG 2100
Db 2041 TCGAAAAACGCTTACGTTGATTTGATTTAGGCGCGGATTTCAAGCGGCTATTTACTTTG 2100
QY 2101 TGGGGCCCAATATATCAGAGAGATTTGTTGAAGCTCTGAATATCATATCTATTTAAGC 2160
Db 2101 TGGGGCCCAATATATCAGAGAGATTTGTTGAAGCTCTGAATATCATATCTATTTAAGC 2160
QY 2161 TAATGACGACGAGGTCGATCTTCTAAAGCATATCCATAGCTTCAAGGGGAATTAAGG 2220
Db 2161 TAATGACGACGAGGTCGATCTTCTAAAGCATATCCATAGCTTCAAGGGGAATTAAGG 2220
QY 2221 AAGCAAAATTAACCGGCTAGCATTTGATTTAGTAAACGAGAAAGTGGAAAGTGGAG 2280
Db 2221 AAGCAAAATTAACCGGCTAGCATTTGATTTAGTAAACGAGAAAGTGGAAAGTGGAG 2280
QY 2281 AAGAGTTGTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2340
Db 2281 AAGAGTTGTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2340
QY 2341 TAATTTTGAAGAAATGTTAGCATTTCTTAAAGCATATCCATAGCTTCAAGGGGAATTAAGG 2400
Db 2341 TAATTTTGAAGAAATGTTAGCATTTCTTAAAGCATATCCATAGCTTCAAGGGGAATTAAGG 2400
QY 2401 TGTGTACGCTTGAATTTTATACGCAACGATGACGGGTTTACTGGAACACGATTC 2460
Db 2401 TGTGTACGCTTGAATTTTATACGCAACGATGACGGGTTTACTGGAACACGATTC 2460
QY 2461 TTGATCTGTGAAGACATCATTTTACACCGGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2520
Db 2461 TTGATCTGTGAAGACATCATTTTACACCGGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2520
QY 2521 AAGGTAATTCGCTTTTAAAGATGATTAATATCATGCTTCTTGAACGGGGTACTGTTG 2580
Db 2521 AAGGTAATTCGCTTTTAAAGATGATTAATATCATGCTTCTTGAACGGGGTACTGTTG 2580
QY 2581 TAGTTGATATAAAGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2640
Db 2581 TAGTTGATATAAAGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2640
QY 2641 TAAGTTATTAAGTTTTCCTTCGTTGTCACACCGGTCGAGGTTCTTGTAAAGCAGTTAT 2700
Db 2641 TAAGTTATTAAGTTTTCCTTCGTTGTCACACCGGTCGAGGTTCTTGTAAAGCAGTTAT 2700
QY 2701 TTTATGTTTATGACGACACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2760
Db 2701 TTTATGTTTATGACGACACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2760
QY 2761 TTGTTCTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAA 2792
Db 2761 TTGTTCTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAA 2792

RESULT 2

ADW98942
ID ADW98942 standard; DNA; 3117 BP.

XX
AC
XX
ADW98942;

DT 01-JUL-2004 (first entry)
XX Diterpene synthase DNA #2.
DE Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;
XX diterpene precursor; diterpene synthase; defence toxin;
KW volatile defensive signal; pollinator attractant; photoprotectant.
KW Stevia rebaudiana.
OS US2004072323-A1.
XX PN 15-APR-2004.
XX PD 07-JAN-2002; 2002US-00041018.
XX PF 05-JAN-2001; 2001US-0259880P.
XX PR (MATS/) MATSUDA S P T.
XX PA (HART/) HART E A.
XX PI Matsuda SPT, Hart EA;
XX DR WPI; 2004-373921/35.
XX PS Claim 3; SEQ ID NO 362; 38pp; English.
XX CC The invention relates to a unicellular organism for producing a diterpene or diterpene precursor comprising an exogenous nucleic acid sequence encoding a geranylgeranyl pyrophosphate synthase under the control of a promoter operable in the organism, and an exogenous nucleic acid sequence encoding a diterpene synthase under the control of a promoter operable in the organism. The invention also relates to methods of producing a diterpene or diterpene precursor and a method of isolating a diterpene synthase comprising growing several cells in the presence of a polyaromatic resin to make a cell/resin mixture, where at least one of the cells further comprises at least one isolated and purified nucleic acid sequence of a yeast expression library, and the expression of the nucleic acid sequence is regulated by an inducible promoter under conditions where the expression is induced, filtering the cell/resin mixture, extracting the cell/resin mixture with alcohol to produce an organic eluent and analysing the organic eluent by a screening method including chromatography and/or spectroscopy, to identify the nucleic acid sequence encoding the diterpene synthase. The unicellular microorganism is useful as a diterpene or diterpene precursor producing system. Diterpenes, in plants, serve as defence toxins, volatile defensive signals, pollinator attractants and photoprotectants. This sequence represents DNA encoding a diterpene synthase polypeptide of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

QY Sequence 3117 BP; 986 A; 502 C; 613 G; 1016 T; 0 U; 0 Other;

Query Match 92.5%; Score 2582.6; DB 12; Length 3117;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 2603; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 140 TTGCGAATTTAGTAATCAGCAAACTAAACATCAATGAATCTTTTCACTATGCTATCGGCTC 199
Db 466 TTGGTGTTCAGTAATCAGCAAACTAAACATCAATGAATCTTTTCACTATGCTATCGGCTC 525
QY 200 CCCTTTGTTAAACCAATCAATCGACCGGGCTCTGTAGCTATTTCATCAGCATCAAC 259
Db 526 CCCTTTGTTAAACCAATCAATCGACCGGGCTCTGTAGCTATTTCATCAGCATCAAC 585
QY 260 TTCACATGGTGGACAACTAAATCCCACTAATCTGATCATTGTATACACCAACCAACGGAT 319
Db 586 TTCACATGGTGGACAACTAATCCCACTAATCTGATCATTGTATACACCAACCAACGGAT 645

QY 320 CCAAAAACAGTTTAAAGATGAGAAATTCGTCTTCATATGACACAGCATGGGTAGC 379
DB 646 CCAAAAACAGTTTAAAGATGAGAAATTCGTCTTCATATGACACAGCATGGGTAGC 705
QY 380 CATGTCCTCTCCAAACTCACCAAAATCGCTTGTTCCTGAGTGTCTCAATTTGGTT 439
DB 706 CATGTCCTCTCCAAACTCACCAAAATCGCTTGTTCCTGAGTGTCTCAATTTGGTT 765
QY 440 AATTAATTAATCAGTTTAAAGATGAGAAATTCGTCTTCATATGACACAGCATGGGTAGC 499
DB 766 AATTAATTAATCAGTTTAAAGATGAGAAATTCGTCTTCATATGACACAGCATGGGTAGC 825
QY 500 TCACCGTTGCTTAAAGATTCCTATCTTCACATTAAGATGAGAAATTCGTCTTCATATGACACAGCATGGGTAGC 559
DB 826 TCACCGTTGCTTAAAGATTCCTATCTTCACATTAAGATGAGAAATTCGTCTTCATATGACACAGCATGGGTAGC 885
QY 560 ATGGAATGTTGGGGAAGATCAAAATAAATAAGATTCCTATCTTCACATTAAGATGAGAAATTCGTCTTCATATGACACAGCATGGGTAGC 619
DB 886 ATGGAATGTTGGGGAAGATCAAAATAAATAAGATTCCTATCTTCACATTAAGATGAGAAATTCGTCTTCATATGACACAGCATGGGTAGC 945
QY 620 TTACGCTACTGAAAAAGATCAACATCTCCATTTGGTTTGGATGACATCAATTTCTGTTT 679
DB 946 TTACGCTACTGAAAAAGATCAACATCTCCATTTGGTTTGGATGACATCAATTTCTGTTT 1005
QY 680 GCTTGAAGTATGCAAAAACTTGACATAAATCTCTTCCAAACAAACAGATTTAGTTT 739
DB 1006 GCTTGAAGTATGCAAAAACTTGACATAAATCTCTTCCAAACAAACAGATTTAGTTT 1065
QY 740 GATGCTACATAAGAGGAATTTGAGCAAAAAAGATCCATTCATTAAGATGAGATGAGATA 799
DB 1066 GATGCTACATAAGAGGAATTTGAGCAAAAAAGATCCATTCATTAAGATGAGATGAGATA 1125
QY 800 CTGCGGTATATCTCTGAAGGACTCGGTAAATTTATATGATGGAATATGATGAGAAATA 859
DB 1126 CTGCGGTATATCTCTGAAGGACTCGGTAAATTTATATGATGGAATATGATGAGAAATA 1185
QY 860 TCAGATGAAGATGTTCTGTTTCACTCACCATCAGCAACAGCTGCTTTCATTA 919
DB 1186 TCAGATGAAGATGTTCTGTTTCACTCACCATCAGCAACAGCTGCTTTCATTA 1245
QY 920 TCATCAAAATCTCTGTTGTTTAAATTTCACTTTGCAAGATTTGTTAATGC 979
DB 1246 TCATCAAAATCTCTGTTGTTTAAATTTCACTTTGCAAGATTTGTTAATGC 1305
QY 980 AGTCCCAACAGTTTATCTCATGATTTATTCGACATTTCTATGTTGACACAAATGA 1039
DB 1306 AGTCCCAACAGTTTATCTCATGATTTATTCGACATTTCTATGTTGACACAAATGA 1365
QY 1040 AAGATTAGGAATTTCAACCATTTCAAGTGGGAAATTTAAATTTAGATGAACATA 1099
DB 1366 AAGATTAGGAATTTCAACCATTTCAAGTGGGAAATTTAAATTTAGATGAACATA 1425
QY 1100 CAGATGTTGGGTGGAACAGATGAGCAAAATTTATGATGATGTTGTAACATGCTTTAGC 1159
DB 1426 CAGATGTTGGGTGGAACAGATGAGCAAAATTTATGATGATGTTGTAACATGCTTTAGC 1485
QY 1160 CTTTCCGTTTAAAGATGAGTAAAGTTTCCGAGATCCATTTGCTGAAATTC 1219
DB 1486 CTTTCCGTTTAAAGATGAGTAAAGTTTCCGAGATCCATTTGCTGAAATTC 1545
QY 1220 TAATGAATTAAGTTTGAAGACGAATGATGAGCTTTGAAACATATCATGCTGACATAT 1279
DB 1546 TAATGAATTAAGTTTGAAGACGAATGATGAGCTTTGAAACATATCATGCTGACATAT 1605
QY 1280 ATTATACCAAGAGATTTATCTTGAAACAAATCTGAGTCAAGTCAATTTCTCAA 1339
DB 1606 ATTATACCAAGAGATTTATCTTGAAACAAATCTGAGTCAAGTCAATTTCTCAA 1665
QY 1340 AGAGATAATATCCACTGATTAACAGGCTTTCTAAATTAATTCACAAAGAGGTGAAAA 1399
DB 1666 AGGATATATATCCACTGATTAACAGGCTTTCTAAATTAATTCACAAAGAGGTGAAAA 1725

QY 1400 TGCTCTTAAGTTCCCTATCAATACCGGTTTGAACGCAATAAACACTAGACGAATATACA 1459
DB 1726 TGCTCTTAAGTTCCCTATCAATACCGGTTTGAACGCAATAAACACTAGACGAATATACA 1785
QY 1460 GCTTTACATATGAGCAATACCAAGATTTCTGAAACTACATATCACTCATCAAAATATTAG 1519
DB 1786 GCTTTACATATGAGCAATACCAAGATTTCTGAAACTACATATCACTCATCAAAATATTAG 1845
QY 1520 TAACTCATTAACCTTAAGTTGGCTTGTGAAGATTTCTACACCTCCCAATCTATTATCG 1579
DB 1846 TAACTCATTAACCTTAAGTTGGCTTGTGAAGATTTCTACACCTCCCAATCTATTATCG 1905
QY 1580 TGAAGAATTAAGGCTTGTGAAGATTTCTGAACTAGAGATTAAGTTGGACCACTCAAGTT 1639
DB 1906 TGAAGAATTAAGGCTTGTGAAGATTTCTGAACTAGAGATTAAGTTGGACCACTCAAGTT 1965
QY 1640 TGCTAGGCAAAAGACCGCTACTGTTTCTGCTGCTCAACACATTTCTGCTCTCCGA 1699
DB 1966 TGCTAGGCAAAAGACCGCTACTGTTTCTGCTGCTCAACACATTTCTGCTCTCCGA 2025
QY 1700 ATTATCAGATGCGCTATTTCTATGCGCCAAATAATGCGATTTAACTACAGTATGATGA 1759
DB 2026 ATTATCAGATGCGCTATTTCTATGCGCCAAATAATGCGATTTAACTACAGTATGATGA 2085
QY 1760 CTTTCTTGAATTCGCTGATCAATCGATGAATTTGACCACTGATTTCAATGTTGAAAA 1819
DB 2086 CTTTCTTGAATTCGCTGATCAATCGATGAATTTGACCACTGATTTCAATGTTGAAAA 2145
QY 1820 ATGGAATGATGATGTCGCAAGGATTTGTTGTCAGAGCATGTTCCGATTTATTTTAGC 1879
DB 2146 ATGGAATGATGATGTCGCAAGGATTTGTTGTCAGAGCATGTTCCGATTTATTTTAGC 2205
QY 1880 ATTAAAAAGATGCAATCTGTTGATTTGGAGATGAAGCTTTTAAATGCAAGCGCGCATGT 1939
DB 2206 ATTAAAAAGATGCAATCTGTTGATTTGGAGATGAAGCTTTTAAATGCAAGCGCGCATGT 2265
QY 1940 AACTAGCCATGTTTCAAACTGTTGAACTTAATGAATAGTATGTTGAGAGAACTAT 1999
DB 2266 AACTAGCCATGTTTCAAACTGTTGAACTTAATGAATAGTATGTTGAGAGAACTAT 2325
QY 2000 ATGGAACAGAGATGCTTATGTCACCAATTTAAATGAATATATGAAAAACGTTTACGTGC 2059
DB 2326 ATGGAACAGAGATGCTTATGTCACCAATTTAAATGAATATATGAAAAACGTTTACGTGC 2385
QY 2060 ATTTGCAATTAGGCCCGCATTTCAAGCGCGCTATTTCTTTTGGGGCCCAAAATATCAGA 2119
DB 2386 ATTTGCAATTAGGCCCGCATTTCAAGCGCGCTATTTCTTTTGGGGCCCAAAATATCAGA 2445
QY 2120 GAGATGTTTGAAGCTCTGAATATCATTAATTTAAGCTAATGAGCAGCAGGCTCG 2179
DB 2446 GAGATGTTTGAAGCTCTGAATATCATTAATTTAAGCTAATGAGCAGCAGGCTCG 2505
QY 2180 ACTTTAAACGATATCCATAGCTTCAAGGGGAATTTAAGGAAGGCAAAATTAACCGCGT 2239
DB 2506 ACTTTAAACGATATCCATAGCTTCAAGGGGAATTTAAGGAAGGCAAAATTAACCGCGT 2565
QY 2240 AGCATTCGATTTGAGTAAACGACAAAGTGGAAAGTGGAAAGACAGGTTGTGAGAGAT 2299
DB 2566 AGCATTCGATTTGAGTAAACGACAAAGTGGAAAGTGGAAAGACAGGTTGTGAGAGAT 2625
QY 2300 GATGATGATTAATAAACAAGAGGAATTAATGAATTAATTTTGAAGAAAAATGG 2359
DB 2626 GATGATGATTAATAAACAAGAGGAATTAATGAATTAATTTTGAAGAAAAATGG 2685
QY 2360 TAGCATTTCTTAGAGCTTGTAAAGATGCAATTTTGAACATGCTGTCACGTGTTGAATTT 2419
DB 2686 TAGCATTTCTTAGAGCTTGTAAAGATGCAATTTTGAACATGCTGTCACGTGTTGAATTT 2745
QY 2420 TTTTACGCAACGATGACGGTTTACTGAAACACAGATTTCTTGATCTGTGAGGACAT 2479
DB 2746 TTTTACGCAACGATGACGGTTTACTGAAACACAGATTTCTTGATCTGTGAGGACAT 2805
QY 2480 CATTTACACCCGTTGCTGTTGATGAAATGAAAGTGAAGAACAAAGGTAATTTGGCTTTAA 2539

Db 2806 CATTTACAAACCGTGTGCTGTGAAATGAAATGAAGAACAAAGTAATTGCGCTTTAA 2865
Qy 2540 AAGATGATAAATAATCATGCTCTTTCACGGGGTACTGTGTAGTTGTATATAAAGGTT 2599
Db 2866 AAGATGATAAATAATCATGCTCTTTCACGGGGTACTGTGTAGTTGTATATAAAGGTT 2925
Qy 2600 GTAGTTGTATATAAAGGTAATAGTAAATCAATAGAACCTTAAGTTAATTAAGTTTTC 2659
Db 2926 GTAGTTGTATATAAAGGTAATAGTAAATCAATAGAACCTTAAGTTAATTAAGTTTTC 2985
Qy 2660 CTCGTGTACACACACGGTGAGGTTCTGTGTAAAGCAGTTTATTTATTTAGTTAGTACGAC 2719
Db 2986 CTCGTGTACACACACGGTGAGGTTCTGTGTAAAGCAGTTTATTTATTTAGTTAGTACGAC 3045
Qy 2720 TTGGTGGTGTGCTGTGTTTATCTCTTAAGAGGCTTTCATTTGTTCTTAAAAAAA 2776
Db 3046 TTGGTGGTGTGCTGTGTTTATCTCTTAAGAGGCTTTCATTTGTTCTTAAATAAGAA 3102

RESULT 3

ADM98948
XX ADM98948 standard; DNA; 2594 BP.
AC
XX
XX
DT 01-JUL-2004 (first entry)
XX
DE Diterpene synthase DNA #8.
XX
XX Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;
KW diterpene precursor; diterpene synthase; defence toxin;
KW volatile defensive signal; pollinator attractant; photoprotectant.
OS Lactuca sativa.
XX
FN US2004072323-A1.
XX
XX 15-APR-2004.
XX
XX 07-JAN-2002; 2002US-00041018.
XX
XX 05-JAN-2001; 2001US-0259880P.
XX
XX (MATS/) MATSUDA S P T.
PA (HART/) HART E A.
XX
XX Matsuda SPT, Hart EA;
PI
XX
XX WPI; 2004-373921/35.
XX
XX New unicellular organisms comprising exogenous nucleic acids encoding a
PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for
PT producing diterpenes and diterpene precursors.
XX
XX Claim 3; SEQ ID NO 368; 38pp; English.

The invention relates to a unicellular organism for producing a diterpene or diterpene precursor comprising an exogenous nucleic acid sequence encoding a geranylgeranyl pyrophosphate synthase under the control of a promoter operable in the organism, and an exogenous nucleic acid sequence encoding a diterpene synthase under the control of a promoter operable in the organism. The invention also relates to methods of producing a diterpene or diterpene precursor and a method of isolating a diterpene synthase comprising growing several cells in the presence of a polyaromatic resin to make a cell/resin mixture, where at least one of the cells further comprises at least one isolated and purified nucleic acid sequence of a yeast expression library, and the expression of the nucleic acid sequence is regulated by an inducible promoter under conditions where the expression is induced, filtering the cell/resin mixture, extracting the cell/resin mixture with alcohol to produce an organic eluent and analysing the organic eluent by a screening method including chromatography and/or spectroscopy, to identify the nucleic

CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents DNA encoding a diterpene synthase polypeptide of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification but was obtained in electronic format from
CC USPTO at seqdata.uspto.gov/sequence.html.

XX
SQ Sequence 2594 BP; 812 A; 444 C; 532 G; 806 T; 0 U; 0 Other;

Query Match 51.2%; Score 1428.6; DB 12; Length 2594;
Best Local Similarity 75.8%; Pred. No. 0;
Matches 1833; Conservative 0; Mismatches 529; Indels 57; Gaps 3;

QY 238 CAGCTATTTCATACAGCACTCACTTCACATGGTGGACAACTAAATCCCACTAATCTGATCA 297
Db 187 CATCGGGTCTTCGTACAGCTTCTTCACAAGCTGGACAAGTTAATCTCTACTGTCATGCC 246
QY 298 TTGATACACCAAGAACGGATCCAAAAACAGTTTAAAAATGTAGAAATTTCTGTTCTT 357
Db 247 TTGATGTGACCAAGAACCAATCCGAAAGCTGTTCACAATGTGGAAGTTTCTGTTCTT 306
QY 358 CATATGACACAGCATGGGTAGCATGCTCCCTTCTCCAACTCACCACAAATCGCTTGT 417
Db 307 CATATGACACAGCTTGGGTAGCCATGGTCCCTTCTCCAACTCCTCCAAATCCCCCTT 366
QY 418 TCCCTGAGTGTCTCAATGGTTAATTAATATCAAGCTTAAATGATGTTTCATGGGTCTTG 477
Db 367 TCCCTGATGTCTGAACCTGTTTACTGGATAATCAAGCTTGAATGATGTTTCATGGGTCT 424
QY 478 TTAATCACACTCAATATCAATATCACCGTGTCTTAAAGATTCTATCTTCAACATTAG 537
Db 425 -----TCTTCCTCATCAGTCCCAATTAATTAAGATACTCTCTCTTCAACATTAG 474
QY 538 CATGTATTGTTGCAATTAAGAGATGGAATGTTGGGGAAGATCAATAATAAAGTCTAA 597
Db 475 CATGTACTTTCATTAAGACATGGAATGTTGGGAAGACCAATTAACAAAGTTTAC 534
QY 598 GTTTTATGATCAAAATCTTCTGCTCAGCTACTGAAAAAAGTCAACCATCTCCCATTTGTT 657
Db 535 ATTACATAGTCAAAATTTTCTGCTCAGTCACTGACAAAAATCAAGCACTCTCCATTGGT 594
QY 658 TTGACATCATATTTCTGCTGTTGTTGATGCTGGAATAAAGTCTTGAACATAAAGTCTTCT 717
Db 595 TTGACATCATATTTCTGCTGTTGTTGATGCTGGAATAAAGTCTTGAATAAAGTCTTCT 654
QY 718 CAAAAACAAACAGATTTTGTATGCTACATAAGAGGAATTCGAGCAAAAAAGATGCC 777
Db 655 TAAACCAACACATTTGAGTGTGATTTCAATGAGAGAAATTCGAGCTAAGAGATGTC 714
QY 778 ATTCAATGAGAGATGGATGATCTGCGTATATCTCTGAAGGACTCGGTAATTTATATG 837
Db 715 ATTCAAATGGAGGAAGCATCTTGCATATATCTCAGAAGGACTTGGAAATTTAAATG 774
QY 838 ATTGGAATATGTCAGCAAAATATCAGATGAAATGTTCTGTTTCAACTCACAATCAG 897
Db 775 ACTGGAACATGGTGATGAAATATCAAAATGAAGATGTTCTCTTTCAACTCACCCTCAG 834
QY 898 CAACAGCTGCTGCTTTCATTAATCATCAAAATCCTGTTGTTGTTAATTTAATTTAC 957
Db 835 CAACAGCTTCTGTTCTTATTCATCAUCAAATGCTGTTGTTCTTCAATATCACTTAC 894
QY 958 TTTTGGCAAGTTTGGTAAATGCAAGTTCACCAAGTTTATCTCATGATTTATTTATCCGAC 1017
Db 895 TCTTTGCAAAATTTGGAATGCAAGTTCACCAAGTTTATCCCTATTGATTTATATGAGGC 954
QY 1018 TTTCTATGTTGACACAAATTAAGATTAAGAAATTTCAACATTTTCAAGTGGAAATTA 1077
Db 955 TTTTCCATGTTGACACACTTTGAAAGATTGGGAATCAAAACGACATTTTCATGTTGAAATCC 1014
QY 1078 AAAATGTTTATGATGAACATACAGATGTTGGGTGGAAACGAGATGACCAATATTCATGG 1137

Db 1015 AAAATGTTTGTAGTAAACATATAGATGTTGGGTGAGGGGATGTGCAAAATATTCATGG 1074
 QY 1138 AFGTTGTAAACATGCTGTTAGCTTTTCGGTTATTAAGATCAATGGGTATGAAGTTTCC 1197
 Db 1075 ATGTTGTACATGTTGTAGCTTTTCGGGTATTAAGAGCAATGGGTATGAAGTCTCTT 1134
 QY 1198 CAGATCCATGCTGCTGAATTTACTAA-----TGAAATGAGCTT 1233
 Db 1135 CGGATCCATGCTGCTGAATTTACTAAAGAGGGGATTAATGATTTCACTGAAAGCCTT 1194
 QY 1234 TGAAGAGCAATATGATGAGCTCTTGAACATATCATGCGTCAATATATATATATATCAAGAGG 1293
 Db 1195 TCAAGAGCCTATACATCTCTTGAAGTTTATAAGGCATCAAGATCATATATCAAGAGG 1254
 QY 1294 ATTTATCTTCTGAAAAAACHAATCTTGAAGTCAAGTCAAGTCTCAAGAGATATATATCA 1353
 Db 1255 AGTTAGCTTTTAGAACAACAACTTGA-----CAAGTTATCTCC 1293
 QY 1354 CTGATTTCAACAGGCTTTCTAAATTAATTCACAAAGAGGTGAAATGCTCTTAAAGTTCC 1413
 Db 1294 CTTCAAGTAAACAGCTCTCTATATATATCTCAAGAGGTGATGATGCTCTTAAAGTTCC 1353
 QY 1414 CTATCAATACCGTTTGAACAGCATATAACATGAGCAATATACAGCTTTTCAATGTAG 1473
 Db 1354 CTTTAAATGCGAGCTAGAACGATGTCCTACTAGAGAAACATAGAGCAATTTCAATTTAA 1413
 QY 1474 ACAATACAGAAATCTGAAACCTTACATATATCACTCAATATATAGTAAACATGATTC 1533
 Db 1414 ACCATAACAAGATCTTTGAAACCTTACATATATAGTCTCAACAACATTTAGCAACAAGATTC 1473
 QY 1534 TAAAGTTGCTGTTGAAGATTTCTACACCTGCAATCTATTTATCTGGAAGATTTAAAG 1593
 Db 1474 TAAAGTTAGCTGCAAGATTTCAATGAATGCCATCTATCTATTTGGAAGTAAAG 1533
 QY 1594 GTCTTGAAGAGTGGGTGAGTAAAGTAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1653
 Db 1534 ATCTTGAAGAGTGGGTGAGTAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1593
 QY 1654 CGGCTACTGTTATTTCTGTTGCTGCAACATTTGCTCCGGAATATATAGATGAGTGGC 1713
 Db 1594 CAGCTACTGTTATTTCTGCTGCTGCTATTTCTTCTCCCTGATTTATCATGATGCCC 1653
 QY 1714 GTATTTTCAAGGGCAAAATATGCAATTAATCACTAGTATGATGATGATGATGATGATGAT 1773
 Db 1654 GAAATTTCAAGGGCAAAATAGTATGATGATGATGATGATGATGATGATGATGATGATG 1713
 QY 1774 GTGGTAAATCGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1833
 Db 1714 GAGGATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1773
 QY 1834 TCGACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1893
 Db 1774 TTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1833
 QY 1894 TCTGTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1953
 Db 1834 TCTGTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1893
 QY 1954 TTCAAACTTGGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2013
 Db 1894 TTGAAATTTGGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1953
 QY 2014 CTTATGTCGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2073
 Db 1954 GTTCAATTTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2013
 QY 2074 CAGATGTCAGCGGCTATTTTACTTTTGGGGGCCCCAAATTTATCAGAGGAGATTTGTTG 2133
 Db 2014 CAGATGTCCTCTTACTCTTTTACTTTTGGGGGTTAAATTTGTCGGAAGAGTGTGTTG 2073
 QY 2134 GCTCTGAATATCATATCTATTTAAGCTATAGCAGCAGGCTGATCTCTTAAACGATA 2193
 Db 2074 GCTCCGAGTATCAAGCTTTATGAGTTTATGAGCTATGAGCACTCAGGCTCGTCTCATGA 2133

QY 2194 TCCATAGCTTCAAGAGGGAATTTAAGGAAGGCAAAATTAACCGGTAGCATTTGCAATTTGA 2253
 Db 2134 TCCATAGCTTCAAGAGGGAATTTAAGGAAGGCGGGGAATTTGAATGCTGTGCAATTTG 2193
 QY 2254 GTACCGGAGAAAGTGGGAAAGTGGGAAAGAGGTTGTGGAGGAGATGATGATGATTA 2313
 Db 2194 GTGATGGAAGAGTGGGAGTGTGGGAAAGAGGTTGTAGAGGAGATGAAGATTTTGACAA 2253
 QY 2314 AAAACAAGGGAAGAAATTAATGAATTAATTTTGAAGAAATGATGATGATTTTCCCTA 2373
 Db 2254 AAAGTCAAGGGAAGAGATGATGAATTTAGTTTGGAAACAAGAGAGTGTGTTCCAA 2313
 QY 2374 GAGCTTGTAAAGATGCAATTTTGGAAACATGTCAGTGTGTCAGTGTGTCAGTGTGTC 2433
 Db 2314 GAGTATGCAAGATGATTTTGGAAACATGTCAGTGTGTCAGTGTGTCAGTGTGTCAG 2373
 QY 2434 ATGACGGGTTTACTGAAACACGATTTCTTGATGATGTCGTCAGGACATCATTTCAACCC 2493
 Db 2374 ATGATGGGTTCACTGGAATGCAATTTCTTGATGATGTCAGGAAATATATATGAACTG 2433
 QY 2494 TGGTCTGTCGATGAATGAAGATGAAGCAAGGTAATTTGGCCCTTTAAAGATGATAATA 2553
 Db 2434 TGTCCATGAGTGTGATATATCAATTTGTCATCAATTCGATATATATATATGATGAT 2493
 QY 2554 ATCATGCTCTTGACGGGGTACTGTTGTAGTTGTATATAATAAGGTTGTAGTTGTATA 2613
 Db 2494 AAAAATGAACAAGCATACTAGTATTTCTAGATCTACTCATGTTTATTTATTTATTT 2553
 QY 2614 AAGGTAATAGGTAATCAAT 2632
 Db 2554 AATATAGTTTGTAGCTAT 2572

RESULT 4

ADM98947
 ID ADM98947 standard; DNA; 2638 BP.
 XX
 AC ADM98947;
 XX
 DT 01-JUL-2004 (first entry)
 XX
 DE Diterpene synthase DNA #7.
 XX
 KW Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;
 KW diterpene precursor; diterpene synthase; defence toxin;
 KW volatile defensive signal; pollinator attractant; photoprotectant.
 XX
 OS Cucumis sativus.
 XX
 PN US2004072323-A1.
 XX
 PD 15-APR-2004.
 XX
 PF 07-JAN-2002; 2002US-00041018.
 XX
 PR 05-JAN-2001; 2001US-0259880P.
 XX
 PA (MATS/) MATSUDA S P T.
 PA (HART/) HART E A.
 XX
 PI Matsuda SPT, Hart EA;
 XX
 XX WPI; 2004-373921/35.
 XX
 PT New unicellular organisms comprising exogenous nucleic acids encoding a
 PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for
 PT producing diterpene and diterpene precursors.
 XX
 PS Claim 3; SEQ ID NO 367; 38pp; English.
 XX
 CC The invention relates to a unicellular organism for producing a diterpene
 CC or diterpene precursor comprising an exogenous nucleic acid sequence

CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
 CC promoter operable in the organism, and an exogenous nucleic acid sequence
 CC encoding a diterpene synthase under the control of a promoter operable in
 CC the organism. The invention also relates to methods of producing a
 CC diterpene or diterpene precursor and a method of isolating a diterpene
 CC synthase comprising growing several cells in the presence of a
 CC polyaromatic resin to make a cell/resin mixture, where at least one of
 CC the cells further comprises at least one isolated and purified nucleic
 CC acid sequence of a yeast expression library, and the expression of the
 CC nucleic acid sequence is regulated by an inducible promoter under
 CC conditions where the expression is induced, filtering the cell/resin
 CC mixture, extracting the cell/resin mixture with alcohol to produce an
 CC organic eluent and analysing the organic eluent by a screening method
 CC including chromatography and/or spectroscopy, to identify the nucleic
 CC acid sequence encoding the diterpene synthase. The unicellular
 CC microorganism is useful as a diterpene or diterpene precursor producing
 CC system. Diterpenes, in plants, serve as defence toxins, volatile
 CC defensive signals, pollinator attractants and photoprotectants. This
 CC sequence represents DNA encoding a diterpene synthase polypeptide of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification but was obtained in electronic format from
 CC USPTO at seqdata.uspto.gov/sequence.html.

XX
 SQ Sequence 2638 BP; 810 A; 466 C; 581 G; 781 T; 0 U; 0 Other;
 Query Match 29.0%; Score 809.4; DB 12; Length 2638;
 Best Local Similarity 62.0%; Pred. No. 1e-180;
 Matches 1410; Conservative 0; Mismatches 811; Indels 54; Gaps 6;

QY	267	GGTGACAACTAATCCCACTAATCTGATCATATGATACAAACCAAGACGGATCAAAAA	326
DB	159	GGCACAAAACTAAACTGGAGCTTTGGCTTTGAAGAAACAAAGAAAGAAATAAAAA	218
QY	327	CAGTTTAAATATGAGAAATTTCTGTTCTTCATATGACACAGCATGGGTAGCCATGTC	386
DB	219	TTGTTCAAAATATGTAACCTTCAATTTTCATATGATGATCTGATGGGTGGCAATGGTC	278
QY	387	CTTTCTCCAACTCAACCAATCGCTTTGTTCTCGCTGAGTCTCAATGGTTAATTAAT	446
DB	279	CCTTCTCCAACTCTCTTAATAAACCTTTTCTGATGTCATAACTGGGTATTAGAT	338
QY	447	AATCAGCTTAATGATGGTTCAATGGGTTCTGTTAATCACTCACTAATAATCAACCG	506
DB	339	CATCAAAACCCGTGATGGTCAATGGG-----CATCTCCATGACCATCAGTTG	386
QY	507	TTGCTTAAAGATTTCTATCTTCAACATAGCATGATTTGTCATTAAGAAAGATGAAT	566
DB	387	GTGATGAAGCCACTCTCTTATCCACATFATGATGTGTCTTACTCTTAAGCCGATGGAT	446
QY	567	GTGCGGAAGATCAATAAATAAAGGTCTAAGTTTATTTAGTCAAAATCTTGTCTCAGCT	626
DB	447	ATCGGTGATGATCATATGACGCAAGCCCTTAGTTTATCAAGTCTAATATAGCTTCAGCT	506
QY	627	ACTGAAAAAGTCAACCATCTCCCATTTGGTTTGAATCATATTTCTGTTGCTTTGAG	686
DB	507	ACTGATGAGATCAACGTTCTCTGTGGGATTTGACATAATTTTCCCTGGTATGATGAG	566
QY	687	TATGCGAAAACTTGGACATAAACCCTCTTCAAAACAAACAGATTTTAGTTGATGCTA	746
DB	567	TATGCTAAAGACTTGAATTTGAATCTACCTTTGGCATCAATGAATGGATGCTTTGGTT	626
QY	747	CATAAGAGGAATTTGGACAAAAAAGAT-----GCCATTCMAATGAGATGGATGATAC	800
DB	627	CAAAAGAAAGATTTGGAGCTTAGAGCTGCTGTAGCACTCTGAGAGGAAAGCCATAT	686
QY	801	TTGCGGTATATCTGTAAGACTCGGTAATTTATGATTTGGAATATGTTGTAAGAAATAT	860
DB	687	TTAGCGTATGTTTCAGAAAGAAATTTGAAATTTACAGGACTGGGAAATGTTCAATGCGAT	746
QY	861	CAGATGAAATATGTTCTGTTTCAACTCACCATCAGCATGCTGCTCTTCATTAAT	920
DB	747	CAAGGAAGAACCGATCACTGTTTAGTTCTCCATCCACCGGCGATGGCTTTTATGCAC	806

QY	921	CATCAAAATCTCGTGTCTTAATTTAATTTCAATTTGGACAAGTTTGGTAATGCA	980
DB	807	AGAAATGATGAGGCTGTGTTTAAATTTACCTTCGCTCAGTCTTACAAAAGTTTCATAGTTCA	866
QY	981	GTCCCAACAGATTTATCCCTCATGATTTAATTTATCCGACTTTCTATGTTTGACACAATTTGAA	1040
DB	867	GTCCCGCAATATATCTCTTGATATATATGCTGTTTACACATGTTGATGACCTTCAA	926
QY	1041	AGATTAGAAATTCACACCATTTCCAGATGGAATTTAAAAATGTTTATAGTGAACAATAC	1100
DB	927	AAACTGGGGAATGATGGCCATTTCAAAGATGAGATTAGAAGTATTAGATGAACAATAC	986
QY	1101	AGATGTTGGGGAACGAGATGACCAATATTTCTGATGTTGTAACATGCTTACCTTTAGCC	1160
DB	987	AGCTGTTGGATGCAAGGGGAGAAACATATTTCTTAGATGCTTCAACTTTGTGCAATGGCC	1046
QY	1161	TTTCGGTTATTAAAGGATCAATGGGTATGAAGTTTCCCGAGATCCATTTGGCTGAAATTAAT	1220
DB	1047	TTCCGGATGTTACGTGTTGAAGGATATGATCTTTCTTCAGATCAATTTGACTCAATTTCTCA	1106
QY	1221	AA-----TGAAATAGCTTTGAAAGACGAATATGACAGCTTTGAA	1259
DB	1107	GAAGGTCTCTTTTCAAATTTGCTCGGAGGACATTTAAAGAGCTTTAGTGCCTCACTAGAG	1166
QY	1260	ACATATCATGCTCACA---TATATTATACCAAGAGGATTTATCTTCTGGAACAAATC	1316
DB	1167	TTATTTAAGGCTCTCCAGATTAATCAATTTATCCGGATGAGTTTATCTCTGGAATAATAAAC	1226
QY	1317	TTGAAGTCAGCTGATTTTCTCAAAGAGATTAATATCCACTG-----ATTCAAACAGG	1367
DB	1227	TCCTTGGACTAGTCTGTTTCTCGAATCATGGAATATCTAGTGTTCAGTTTCATTTCTGATAGA	1286
QY	1368	CTTCTTAAATTAATTCACAAAGAGGTGGAATAATGCTCTTTAAGTTCCCTATCAATACCGGT	1427
DB	1287	ACTGAGAGACTCGTGAACCAAGAGGAGTAAATGCTTTCGAGTTTCCCTATAATTTCACT	1346
QY	1428	TTGAAACGATAAACACTAGACGGAATATACAGCTTTTACATGTAGATGACAAATACAGAAAT	1487
DB	1347	CTGAAACGCTTATCAAAATTAAGCGGACACTGGAAAGTTACAGTGGAGACATTTGTGAGGAT	1406
QY	1488	CTGAAACTACATATCACTCATCAATATTAGTAAACATGATTAACCTTAAGTTTGGCTGTT	1547
DB	1407	TCAAAACAGCATATGCTGTTTAAATTTTGGTTCATCAAGATTTCTTGGAACTTTGCTGTA	1466
QY	1548	GAAGATTTTACACTGCTCCCAATCTATTTATGCTGAAGAAATTTAAAGCTTTGAAAGGTGG	1607
DB	1467	GAAGATTTCAATACCTTGCAGGCATACATCGCAAGGAACTGAAAGAGCTTGAATAATGG	1526
QY	1608	GTGTTAGAGAAATAGTTGGACCAAGCTCAAGTTTCTAGGCAAAAGACCGCTTACTGTAT	1667
DB	1527	GTATCGAAACAAATTTGGAAGTTGAAATTTGCGAGACAGAAATTTAGCGTACTGCTAT	1586
QY	1668	TTCTCTGTTCTGCAACACTTTGCTCTCCGAATTTATCAGATGCGCTATTTTCATGGGCC	1727
DB	1587	TTTTCTGCTGACGAGCCCTTAACCTCTCAGAACTTTTGTGATGCCCTTATCATGGGCA	1646
QY	1728	AAAAATGGCAATTAATCTACAGTATGATGACTTTTTTGTGATATGCGGTGGTACAAATCGAT	1787
DB	1647	AAAAATGGGTACTCAACACCGGTGTTGATGATTTCTTGTGATGAGGATCTGGAAGAG	1706
QY	1788	GAATTCAGCAACCTGATTCATGTTGAAATTTGAAATTTAGATCTGATGCTGACAGGATTTGT	1847
DB	1707	GAATTTGTAACCTTATACAAATTTGGTGAAGAAATGGGATGCCAGTGGGAAACCGGTATC	1766
QY	1848	TGTTTCAGAGCATGTTTCGGATTTTATTTTATAGCAATTTAAAGATGCAATCTGTTGATTTGA	1907
DB	1767	TGTTTCCAGAGGTTGAGATTTATTTCTTGGCACTTCATAGCAATTTTGGAAATAGGA	1826
QY	1908	GATGAAGCTTTTAAATGGCAAGCGCGATGTAACCTAGCCATGTTATTTCAAACTTGGTTG	1967
DB	1827	AAAAAGCTTTTACCTTGGCAAGGACGACGCTGATGAGGAATGTTATCGATATTTGGTTG	1886
QY	1968	GAATTAATGATATGTTTGAAGAGCTATATGACAAAGATGCTTATGTTGCGCAACA	2027

1887 GCTTGTCTGAGTCAATGAGGAGAACTGAATGTTGAAAAAFAAGTAGTGCATCA 1946
2028 TTAATGAATATATGAAAGAAAGCTTACGTGTCATTTGATTTAGGCCCGGATTTGAAGCG 2087
1947 TTGGATGAGTACATGAAATGGCTATGTATCATTTGCTTTGGGACCTATAGTCTTCCA 2006
2088 GCTATTTACTTTGTTGGGCCCAATATATCAGAGGAGATTTGTTGAAAGCTCTGAATATCAT 2147
2007 AGCTCTTACTTTGTTGGACCTAAGCTTCCAGAGGAAATTTGTTGAAATTTGAAATACAG 2066
2148 AATCTATTTAAGCTAATGAGCAGCAGGCTGCACTTCTTAAACGATATCCATGTTCAAG 2207
2067 AGCTCTTACTTTGTTGGACCTTCTGCGCCCTTCTGAATGATCTGAACTTTTGTAT 2126
2208 AGGAAATTTAAGAGAGGCAATTAATAACGCGGTAGCATTTGCATTTGAGTACGAGAAAT 2267
2127 AGAGAGTCCAGCGAAGGAAATTAATGCTTTGTCTCTATACATGATTAGTCCCGTGT 2186
2268 GGGAAAGTGGAAAGAGAGGTTGTGGAGGAGATGATGATGATTAAACAAAGAGGAAA 2327
2187 AAGCTCACCAAGAGAGGCCACTGAGCAATGAAAGGAGATGTTGATAGACGAGAGA 2246
2328 GAATTAATGAATTAATTTTGAAGAAATGTTAGCATTTGTTCTTAGAGCTTTGTAAGAT 2387
2247 GAATGTTGAGATTAGTTTTCAGGAGAAACAGTACAA---TTCCAAGAGCTTTGTAAGGAT 2303
2388 GCATTTGGAACATGTGTACAGTGTGATTTTTCAGCAACAGTACAGGCTTTACT 2447
2304 TTGTTCTGGAATAGCTGTGTGTGCACTATTTTACAGGAAAGATGATGGGTTTACA 2363
2448 GGAACACAGATCTTGATCTGTAAGGACATCATTTTACACCCGTTGGTGTG 2502
2364 TCCCATGAGTTGATGAACTCTCGGAAAGCTTATTGAAACCCATGTTCTGG 2418

RESULT 5
ADM98958
ID ADM98958 standard; DNA; 2658 BP.
XX ADM98958;
AC ADM98958;
AT 01-JUL-2004 (first entry)
CT Diterpene synthase DNA #18.
DE Unidentified.
XX Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;
KW diterpene precursor; diterpene synthase; defence toxin;
KW volatile defensive signal; pollinator attractant; photoprotectant.
XX Unidentified.
OS US2004072323-A1.
XX 15-APR-2004.
XX 07-JAN-2002; 2002US-00041018.
XX 05-JAN-2001; 2001US-0259880P.
XX (MATS/) MATSUDA S P T.
XX (HART/) HART E A.
XX Matsuda SPT, Hart EA;
XX WPI; 2004-373921/35.
XX New unicellular organisms comprising exogenous nucleic acids encoding a
PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for
PT producing diterpenes and diterpene precursors.
XX Claim 3; SEQ ID NO 378; 38pp; English.
XX

CC The invention relates to a unicellular organism for producing a diterpene
CC or diterpene precursor comprising an exogenous nucleic acid sequence
CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC promoter operable in the organism, and an exogenous nucleic acid sequence
CC encoding a diterpene synthase under the control of a promoter operable in
CC the organism. The invention also relates to methods of producing a
CC diterpene or diterpene precursor and a method of isolating a diterpene
CC synthase comprising growing several cells in the presence of a
CC polyaromatic resin to make a cell/resin mixture, where at least one of
CC the cells further comprises at least one isolated and purified nucleic
CC acid sequence of a yeast expression library, and the expression of the
CC nucleic acid sequence is regulated by an inducible promoter under
CC conditions where the expression is induced, filtering the cell/resin
CC mixture, extracting the cell/resin mixture with alcohol to produce an
CC organic eluent and analysing the organic eluent by a screening method
CC including chromatography and/or spectroscopy, to identify the nucleic
CC acid sequence encoding the diterpene synthase. The unicellular
CC microorganism is useful as a diterpene or diterpene precursor producing
CC system. Diterpenes, in plants, serve as defence toxins, volatile
CC defensive signals, pollinator attractants and photoprotectants. This
CC sequence represents DNA encoding a diterpene synthase polypeptide of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification but was obtained in electronic format from
CC USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 2658 BP; 757 A; 486 C; 626 G; 789 T; 0 U; 0 Other;
Query Match 28.0%; Score 782.2; DB 12; Length 2658;
Best Local Similarity 61.3%; Pred. No. 2.7e-174;
Matches 1390; Conservative 0; Mismatches 823; Indels 54; Gaps 6;
XX
QY 275 AACTAATCCCACTAATCTGATCATATACAAACCAAGAACCGATCCAAACAGTTTAA 334
Db 236 AACTAATACTGGAGCTTTGACACTTTGAAGAAACAAAGAAATTAAGAAATTTGTCGA 295
QY 335 AATGTAGAAATTTCTGTTTCTTATATGACACAGATGGGTAGCCATGCTCCCTCTCC 394
Db 296 CAAGGTTGAACCTTCAGTTTCTGTCATATGATGATGATGATGATGATGATGATGAT 355
QY 395 AACTCAACCAAAATCGCCCTGTTTCCCTGAGTGTCTCAATTTGTTTAAATTAATCAGCT 454
Db 356 AACTCTCTCAACCAACCTCTTTTCCCGAGTGATATAACTGGTATTAGATGATCAACA 415
QY 455 TAATGATGTTTATGAGGCTTTGTTTAAATCACTCATATATATATATATATATATATAT 514
Db 416 TCGTATGAGTCTATGAGGCTTACTTCCACAACTCTCTCTCTCTCTCTCTCTCTCTCT 463
QY 515 AGATTTCTTATCTTCAACATTTAGCATGTTATTTGTCATTTAAAGATGAAATCTTGGGA 574
Db 464 GGCCATCTCTTATCTACATTTAGCATGTTTCTTACTTTAAGCGGTGGAATTTGGGCA 523
QY 575 AGATCAATATAATAAAGGCTAAAGTTTATGAGTCAAAATCTTCTCTCTCTCTCTCTCTCT 634
Db 524 TGAATATATGAGCAAGGCCCTTGAATTTATCAAGTCTATATAGCTTACGATGATGA 583
QY 635 AAGTCAACCATCTCCCATTTGTTTGAATATATTTCTGTTTCTGTTTCTGTTTCTGTTTCT 694
Db 584 GAACCAACGTTCTCCGTTGGGATTTGACATTTATTTTCCCTGGCATGATTGAGTATCTTA 643
QY 695 AACTTTGGACATAAACCTCTTTCAAAACAAACAGATTTTATGTTGATGCTCATATAGAG 754
Db 644 AGACTTTGAATTTGAATCTACCTTTGGCACCGAGAACGTTGGATGCTTGTTCGAAGAA 703
QY 755 GGAATTTGGAGCAAAAA-----AGATGCCATTTCAAAATGAGATGATGATGATGATGATG 808
Db 704 AGAGTTGGAGCTCAGAGCTGCGAAGCAACTCTGAAGGTGGAAAGCCCTATTATTAGCGTA 763
QY 809 TATCTCTGAAGCTCGGTAATTTATGATTTGGAATATGCTCAAGAAATATCAGATGAA 868
Db 764 TGTTCAGAGGAAATTTGGAAGTTACAGGACTGGGATATGTTGATGATGATGATGATGATG 823
QY 869 AATGTTCTGTTTTCAACTCAGCAATCAGCAACAGCTGCTGCTTCTTCAATTAATCATCAAAA 928
XX

Claim 3; SEQ ID NO 381; 38pp; English.

The invention relates to a unicellular organism for producing a diterpene or diterpene precursor comprising an exogenous nucleic acid sequence encoding a geranylgeranyl pyrophosphate synthase under the control of a promoter operable in the organism, and an exogenous nucleic acid sequence encoding a diterpene synthase under the control of a promoter operable in the organism. The invention also relates to methods of producing a diterpene or diterpene precursor and a method of isolating a diterpene synthase comprising growing several cells in the presence of a polyxromatic resin to make a cell/resin mixture, where at least one of the cells further comprises at least one isolated and purified nucleic acid sequence of a yeast expression library, and the expression of the nucleic acid sequence is regulated by an inducible promoter under conditions where the expression is induced, filtering the cell/resin mixture, extracting the cell/resin mixture with alcohol to produce an organic eluent and analysing the organic eluent by a screening method including chromatography and/or spectroscopy, to identify the nucleic acid sequence encoding the diterpene synthase. The unicellular microorganism is useful as a terpene or diterpene precursor producing system. Diterpenes, in plants, serve as defence toxins, volatile defensive signals, pollinator attractants and photoprotectants. This sequence represents DNA encoding a diterpene synthase polypeptide of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from [uspto at segdata.uspto.gov/sequence.html](http://uspto.gov/sequence.html).

Sequence 2658 BP; 757 A; 486 C; 626 G; 789 T; 0 U; 0 Other;

Query Match 28.0%; Score 782.2; DB 12; Length 2658;

Best Local Similarity 61.3%; Pred. No. 2.7e-174;

Matches 1390; Conservative 0; Mismatches 823; Indels 54; Gaps 6;

QY	275	AAC	TAA	TCC	CCCA	CTA	ATCT	GTG	ATCAT	GATAC	ACCA	ACG	GAAC	CGG	ATCC	AAAA	ACAG	TTTAA	333
DB	236	AAC	TAAA	CTG	GAG	CTT	TGCA	CTTT	GA	GA	CA	CA	AA	GA	GA	AA	TAT	GA	295
QY	335	AAA	CTG	AGAA	TTTT	CTG	TTT	CTT	CA	TAT	GACA	CAG	CAT	GGG	TAG	CCAT	TGGT	CCCT	394
DB	296	CA	AGG	T	GAA	CTT	CAG	TTT	CTG	CAT	GAT	ACT	GAT	GCG	TGG	CAAT	TGGT	CCCT	355
QY	395	AAA	CTC	ACCA	AAAT	CGC	CTT	GGT	TTT	CCG	TAG	GTCT	CA	AT	TGG	TTAA	TAT	TAT	454
DB	356	AAA	CT	CT	CT	CA	CC	AA	CC	TCT	TTT	CCG	AG	TG	TATA	AACT	GGG	TAT	415
QY	455	TAA	T	GAT	TGG	TCT	CA	TGG	GGT	CTT	GT	TAA	T	CA	CAC	ACT	CAT	TAA	514
DB	416	TG	CT	GAT	TGG	CT	CAT	TGG	GGC	TACT	CC	CA	CA	AC	-----	GAT	CAG	TTC	463
QY	515	AGA	TT	CT	CT	AT	CTT	CA	AA	CTAG	CAT	GTAT	TG	TG	CATT	AAAA	GAT	GGAT	574
DB	464	G	CC	AA	T	CT	TAT	CTA	CA	ATTAG	CAT	GTG	TTT	CTT	ACT	TAA	GCG	TGG	523
QY	575	AGA	T	CA	AA	T	AAA	T	AA	AG	GT	TTT	AT	T	CAG	T	CAA	AT	634
DB	524	TG	AT	CAT	AT	G	AC	AA	GG	CC	CTT	TGA	TTTT	TAT	CA	AGT	CTA	TAT	583
QY	635	AA	G	T	CA	A	CC	AT	CT	CC	CA	T	GG	TTT	TG	CAT	CAT	AT	694
DB	584	GAA	CA	AA	CG	TT	CT	CG	TG	GG	AT	TG	AC	AT	TTT	CC	TG	CAT	643
QY	695	AAA	CT	TG	G	AC	TAA	AA	CC	CT	CTT	T	CA	AAA	CA	AA	CA	AG	754
DB	644	AG	AT	T	GAA	TTT	TGA	AT	CTA	CC	TT	TG	CA	CG	CA	AC	G	TG	703
QY	755	GG	AA	T	TG	G	AG	CA	AAAA	-----	AG	T	G	CC	AT	T	CA	AT	808
DB	704	AG	AG	T	TG	G	AG	CT	CAG	AA	GT	CG	CA	AG	CA	CA	CT	G	763
QY	809	TAT	CT	CT	GA	AG	CA	CT	CG	GT	AA	TTTT	TAT	TAT	TG	AA	TAT	TG	868
DB	764	TG	TTT	CAG	AA	GG	AA	TT	TG	AA	AG	TTT	TAC	GAG	CT	TGG	AT	TGG	823

Db	884	TCATGGCTGTTTGAATATCTTCGGCTCACTCTTACAAAGTTTGAAGCTCAGTCCAC	943	1964	GAAGTGTATGAGAAAGGAGCTGAATGCTGACAAATAAGGTAGTCCATCAATGGGTGA	2023
Qy	989	AGTTTATCTCATGATTTTATTTATCCGACTTTCTATGGTTGACAAATTGAAGATTAGG	1048	2036	ATATATGAAACAGCTTACGTGTCTATTTGCAATTTAGSCCCGATTTGTCAAGCCGGCTATTTA	2095
Db	944	AAATATCTCTTGATATATATCTCGATACACATGGTTGATAGCTTCAAAATTCGG	1003	2024	ATATATGAAACAGCCCATGTATCATTCGGTGTGGACCTATTAATCTTCCAATGCTCTT	2083
Qy	1049	AAATTCACACATTCAGAGTGGAAATTTAAATGTTTGTAGATGAAACATACAGATGTTG	1108	2096	CTTTGTGGGGCCCAATTTATCAGAGGAGATTGTTGAAAAGCTCTGAATATCATATCTATT	2155
Db	1004	AAATTCCTGGCAATTCAGAGAGAGATTAGAAGCGTATTAGATGAACCTTACAGCTGTTG	1063	2084	CTTTGTGGACCTAACTCTCAGAGGAAATGATTGGAAGCTGTGAATCCAGAGTTATA	2143
Qy	1109	GGTGAACGAGATGACAAATATTCATGATGTGTGAACATGCTGTTTAGCTCTTCGGTT	1168	2156	TTAGCTAATGACACGCGAGTCTTCTTAAACGATATCCATAGCTTCAAGAGGGAATT	2215
Db	1064	GATCAAGGAGAGGAAATATATTTCTAGATGCTTCACTGTCATGTCGCTTTCGAT	1123	2144	TTAGCTGTATGAGCACTGCTGGTGGCTTAAAGAAATGATATTCGATCTTACGATAGAGATG	2203
Qy	1169	ATTAAGGATCAATGGGTATGAAGTTTCCCGAGATCCATTTGGCTGAAATTAATA	1222	2216	TTAGGAAGGCAAAATTAACGCGGTAGCATTTGCAATTTGAGTAACGGGAGAAAGTGGAAAGT	2275
Db	1124	GTTCAGTGTGAAGGATATGATGTTTCTTCAGACCAAGTTTGAATTTTCAGAGATAT	1183	2204	CAAGAGGGAAGCTGAATTTCTCTCTGTGGATGATTGATGGCGGTGTAATGTCAC	2263
Qy	1223	-----TGAATTAGCTTTGAAAGAGATATGATGAGCTCTTTGAAACATATCA	1267	2276	GGAGAAGAGGTTGTGGAGGAGATGATGATGATGATTAATAAACAAGAGAAAGATTAAT	2335
Db	1184	CTTTCCCAATTCCTTTGGAGGATTTTAAAGACTTTCGGTGCCTCGCTGGAGTTATATA	1243	2264	CAAGAGGAGGCCAATTGAAGCAATTAAGGGGATTTTGAGGGCGCATTAAGAGAGCTGCT	2323
Qy	1268	TCGTCACATATATATACCAAGAGGATTTATCTTCT---GGAAAAACAATCTTGAAGTC	1324	2336	GAAATTAATTTTGAAGAAATGGTAGCATTTGTTCTAGAGCTTTGAAAAGATGCAATTTG	2395
Db	1244	GGCTCTCAGATATACCGCACCCGATGATCTGTTCTGGAATATATAAATCTTTGGAC	1303	2324	GGGTTAGTTTTCAGAGGAGACACTACAA---TTCCAAGAGCTTTGTAAGATTTGTTCTG	2380
Qy	1325	AGCTGATTTCTCAAGAGATATATCCACTGANTCAAAAGAGCTTTCTTAA-----	1376	2396	GACATGTCTACGCTGTTCAATTTTTTTTACGCAACGATGACGGGTTTACTGGAACAC	2455
Db	1304	TAGTCGTTTCTGAAGCATGATATCTAGTATTCAGTTTGGTCTGATAGAACCGATAG	1363	2381	GAAATGTATGTCATTTGTAATCTATTTTACATGGAAGATGATGGGTACACTTCAAAATAG	2440
Qy	1377	-TTAATTCACAAAGAGTGAAATGCTCTTAAGTTCCCTATCAATACCGTTTGAAGC	1435	2456	GATTCCTGTACTGTGAAGACATTTTACACCCGTTGGTCTG	2502
Db	1364	TGTTGTTAAACAAGAGGCTTTAATCCTCTGAGTTCCCTTATATGTAACACTCTAGAGGATTT	1423	2441	GTTCATGAACTGTAAAGCCATGTTTGAACCAACCCATGGATCTCG	2487
Qy	1436	CATAACACTAGACGAATATACAGTTTACAATGTAGACAAATACAAAGATTTCTGAAAC	1495	RESULT 8		
Db	1424	CCTAATAAGTAAGAGGCAATGAAAGTTACAGTGGAGACATTTGTGAGGATTTCAAAATC	1483	ADM98957		
Qy	1496	TACATATCACTCATCAATATTAAGTAACACTGATTAACCTAAGTTGGCTGTTGAAGATTT	1555	ID	ADM98957 standard; DNA; 2506 BP.	
Db	1484	GCCATATGCTGCTTAAATTTTGGCCATCAAGATTTTCTGGAACTTCTGTAGAGGATTT	1543	XX	ADM98957;	
Qy	1556	CTACACTGCAATCTATTTATCTGTAAGATTTAAAGGCTTTGAAAGTGGGTGGTGA	1615	AC	01-JUL-2004 (first entry)	
Db	1544	CAATACCTGCAACGCAATTCATCTTAAAGAACTGGAAGAGCTTCAAAGATGGGTGGTGA	1603	XX	Diterpene synthase DNA #17.	
Qy	1616	GAATAAGTTGGACAGCTCAAGTTTGTAGGCAAAAGACCGCTTACTGTTATTTCTCTGT	1675	XX	Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;	
Db	1604	AAACAAATTTGACAGATGAAATTTTTCAGACTGCACTAGGCTACTGCTATTTTGTCTGC	1663	XX	diterpene precursor; diterpene synthase; defence toxin;	
Qy	1676	TGCTGCAACACTTTTCGTTCCGCAATATACAGATGCGCGTATTTTCATGCGGCAAAATGG	1735	XX	volatile defensive signal; pollinator attractant; photoprotectant.	
Db	1664	GGCAGCGACCTTACTGATCTGAACTTCATGATGCTCGCATAGCATGGGCACAAAATGG	1723	OS	Arabidopsis thaliana.	
Qy	1736	CATATTAACACTAGTGTGATGATCTTTTGTATCGGTGGTGAATCGATGAATTCAC	1795	XX	US2004072323-A1.	
Db	1724	TGTCCTCAGCACCGTGTGATGATTTCTATGATGGTGGAGATCTGAAGAGGAATTTGA	1783	XX	15-APR-2004.	
Qy	1796	CAACCTGATTCATGTTGAAAAATGGAATGTAGATGTCACAAAGGATTTCTGTTCCAGA	1855	XX	07-JAN-2002; 2002US-00041018.	
Db	1784	TAACTTATAGAAATTTGTGNAAGTGGATCCTGATGGGGAAGTGGGTACTGTCCAA	1843	XX	05-JAN-2001; 2001US-0259880P.	
Qy	1856	GCATGTTCCGATTTTATTTTATGATTAAGATGCAATCTGTGGATTTGAGATGAGC	1915	XX	(MATS/) MATSUDA S P T.	
Db	1844	GGAGTTGAGATTTGATTTCTTGCACTGCACAGCACAGTTTGTGAAAATAGGAAGAGC	1903	PA	(HART/) HART E A.	
Qy	1916	TTTTAAATGGCAGCGCGATGTAACCTAGCCATGTTATTCAACTGTTGGTGGACTAAT	1975	XX	Matsuda SPT, Hart EA;	
Db	1904	TTTAGTATGGCAAGCAGCGTGTATGAGGAATTTATCGATGGTGTGGTGGCTCTGCT	1963	XX	WPI; 2004-373921/35.	
Qy	1976	GAATAGTATGTTGAGAGAGCTATATGACAAAGAGATGCTTATGTGCCAACATTAATGA	2035	XX	New unicellular organisms comprising exogenous nucleic acids encoding a	
				XX	geranylgeranyl pyrophosphate and a diterpene synthase, useful for	
				XX	producing diterpenes and diterpene precursors.	
				PS	Claim 3; SEQ ID NO 377; 38pp; English.	
				CC	The invention relates to a unicellular organism for producing a diterpene	

or diterpene precursor comprising an exogenous nucleic acid sequence encoding a geranylgeranyl pyrophosphate synthase under the control of a promoter operable in the organism, and an exogenous nucleic acid sequence encoding a diterpene synthase under the control of a promoter operable in the organism. The invention also relates to methods of producing a diterpene or diterpene precursor and a method of isolating a diterpene synthase comprising growing several cells in the presence of a polyaromatic resin to make a cell/resin mixture, where at least one of the cells further comprises at least one isolated and purified nucleic acid sequence of a yeast expression library, and the expression of the nucleic acid sequence is regulated by an inducible promoter under conditions where the expression is induced, filtering the cell/resin mixture, extracting the cell/resin mixture with alcohol to produce an organic eluent and analysing the organic eluent by a screening method including chromatography and/or spectroscopy, to identify the nucleic acid sequence encoding the diterpene synthase. The unicellular microorganism is useful as a diterpene or diterpene precursor producing system. Diterpenes, in plants, serve as defence toxins, volatile defensive signals, pollinator attractants and photoprotectants. This sequence represents DNA encoding a diterpene synthase polypeptide of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

Sequence 2506 BP; 760 A; 454 C; 586 G; 706 T; 0 U; 0 Other;

Query Match 26.2%; Score 732.4; DB 12; Length 2506;
Best Local Similarity 59.5%; Pred. No. 1.5e-162;
Matches 1356; Conservative 0; Mismatches 876; Indels 48; Gaps 5;

QY	270	GGACAACTAATCCCACTAATCTGATCATATTGATACAAACCAAGACGGATCCAAACACAG	329
DB	143	GTACAGACAGAGCTAACAATGTGAGCTTTGACCAACAAAGACAGATTAGGAAGATG	202
QY	330	TTTAAATATGAGAAATTTCTGTTTCTTCATATGACAGAGTGGGTAGCCATGTCCT	389
DB	203	TTGGAAGAGTGGAGCTTCTGTTTTCGGCCTACGATCTAGTTGGGTAGCAATGTTCCA	262
QY	390	TCTCCAACTACCCAAATCGCTTGTTCCTCGAGTGTCTCAATTTGGTTAATTAAT	449
DB	263	TACCGAGCTCCAAATGCTCCACTTTTCCACAGTGTGTGAATGGTTATGGATAAT	322
QY	450	CAGCTTAATGATGGTTCTATGGGTCTTGTATATACATCACTCAATPATAATCACCCGTTG	509
DB	323	CAACATGAAGATGGATCTTGGGAGCTTGATAACC-----ATGACCACTCAATCTCT	373
QY	510	CTTAAAGATTTCTATCTTCAACATATGACATGATTTGTGATTAATAAGATGGAATGTT	569
DB	374	AAGAAGATGTTTATCACTACACTGGCTAGTATPCTTCGGTTTAAAGAGTGGGGAAT	433
QY	570	GGGAAGATCAATAATAAAGGTCTAAGTTTATTGAGTCAAACTCTGCTTCAGCTACT	629
DB	434	GGTGAAGACAAATAACAGGGTCTCCAGTTTATGAGCTGAATTCGCAATTAGTCACT	493
QY	630	GAATAAGTCAACATCTCCATTTGGTTTGGATCATATTTCTCGTTTCTTGGAT	689
DB	494	GATGAACCAATACAGAAACCAACAGGGTTTCAATATATTTCTCGGATGATTAATAT	553
QY	690	GCAGAAACCTTGGACATAAACCCTCTTCAAAACAAACAGATTTTGTGATGCTCAT	749
DB	554	CTTAGATTTGATCTGACGATTTCCATTTGGCTCAGAAAGTGGTGGATGACATATACGA	613
QY	750	AAGAGGAATTTGGAGCAAAAAG-----ATGCCATTCAAAATGAGATGGATGATAC	800
DB	614	AAAAGAGATCTGGATCTTAATGTGATAGTGAAGAAAGTTTCAAAGGGAAGAGAGCATAT	673
QY	801	TTGGCTATATCTCTGAAGGACTCGGTAATTAATATGATTTGGAATATGGTCAAGAAATAT	860
DB	674	CTGGCTATGTTTATAGAGGGGACAAAGAAACCTAAAGATTTGGATTTGATGCAATAT	733
QY	861	CAGATGAATAATGTTCTGTTTTCATCTCACCATCAGCAACAGCTGCTGTTTCAATTAAT	920
DB	734	CAAGGAAATATGGGTCACTGTTTGTATTTCTCCAGCCAAACAGCAGCTGCTTTACTCAG	793

QY	921	CATCAAAATCCGTGTTGTCTTAATTAATTAATAATCTTTTGGACAAGTTTGGTAATGCA	980
DB	794	TTTGGGAATAGTGGTTGTCTCCGTTATCTCTGTTCTCTCTCAGAAATTCGAGGCTGCA	853
QY	981	GTCCCAACAGTTTATCCTCATGATTTATTTATCCGACTTTCATGTTGATGACACAATGAA	1040
DB	854	GTTCCTTCAGTTTATCCATTTGATCAATATGACGCTTAGTATAATTTGCTACTCTTGAA	913
QY	1041	AGATTAGGAATTTCAACACCAATTTCAAGGTGGAAATTAATAATGTTTATAGTGAACATAC	1100
DB	914	AGCTTAGGAATTTGATAGAGATTTCAAAACCGAAATCAAAAGCATATTTGATGAACCTAT	973
QY	1101	AGATGTTGGTGGACAGAGATGAGCAATATTCATGATGTTTAACTATGTTGCTGAAATTACT	1160
DB	974	AGATATGGCTTCGTGGGGATGAAGAAATATGTTTGGACTTGGCCACTTGTGCTTGGCT	1033
QY	1161	TTTCGGTTATTAAGGATCAATCGGTATGAGTTTCCCAAGATCCATGCTGCTGAAATTACT	1220
DB	1034	TTCCGATTAATGCTTGTCTCATGGCTATGATGTTCTTAGATCCGCTAAACCATTTGCA	1093
QY	1221	AATGAATTAGCTTTTGAAGACGAAATATGATGAGCTTTGAAACATATCATCGCTCACATATA	1280
DB	1094	GAAGAATCTGGTTTCTCTGATACTTTTGAAGATATGTTAAGAAACGTTTCTGTGTTA	1153
QY	1281	TTATACCAAGAGGAT-----TTATCTCTGGAACAAATCTTGAAGTCAGCTGATTTCC	1335
DB	1154	GAAATTTATTAAGGCTGTCTCAAGTTTATCCATGAATCAGCTTTGAAGACGAGTGTGT	1213
QY	1336	TCMAAGATATATATCCACTGATTTCAA-----ACAGGCTTCT	1373
DB	1214	TGGACTAAACAAATCTCGAGATGGAATTTGCCAGCTGGGTTAAGACCTCTGTTGAGAT	1273
QY	1374	AAATTAATTTCAAGAGGTGGAAAAATGCTCTTAAGTTTCCCTATCAATACCGTTTAGAA	1433
DB	1274	AAATACCTCAAGAAAGAGTTCGAGATGCTCTTGTCTTCCCTCTCTATGCAAGCCTAGAA	1333
QY	1434	CGCATAAACACTAGACGAAATATATAC---AGCTTTACAATGTAGACAATACAGAAATCTG	1490
DB	1334	AGATCAGATCAACAGAGAAAAATATCTCAATGGTTCTGCTGTGAAAAACACCAAGATTACA	1393
QY	1491	AAAACTACATATCACTCATCAATATTAGTAACACTGATTAACCTTAAGTTTGGCTTTGAA	1550
DB	1394	AAAACCTCATATCGTTTGCAATATTTTGACCTCTGATATCTCTGAAGTTAGCTGTGGAT	1453
QY	1551	GATTTTACACCTCCCAATCTTATTTATCGTGAAGAAATTAAGAGTCTTGAAGGTGGGTG	1610
DB	1454	GACTTCAATTTCTGCCAGTCCATACACCGTGAAGAAATGGAACGCTCTTGATAGGTGGAT	1513
QY	1611	GTAGAGAAATAGTTGGACAGCTCAAGTTTGTAGGCAAAAGACCGCTACTGTTATTTTC	1670
DB	1514	GTGAGAAATAGATTCAGGAACTGAAATTTGCCAGACAGAGCTGGCTTACTGTTATTTTC	1573
QY	1671	TCTGTTCTGCAACACTTTCTGCTCCCGAATTTATCAGATGCGGCTATTTCTATGGGCCAAA	1730
DB	1574	TCTGGGCTGCACTTTATTTTCTCCAGAACTATCTGATGCTCGTATATCTGTTGGGCCAAA	1633
QY	1731	AATGGCATATTAACACTACAGTAGTTGATGACTTTTTCATATCGGTGCTGCAATCGATGAA	1790
DB	1634	GGTGGAGTACTTACAACGGTTGTAGACGACTTCTTTGATGTTGGAGGGTCCAAAGAAAGAA	1693
QY	1791	TTGACCAACCTGATTCATGTTGTGAAAAATGGAATGTAGATGTGCAACAGGATTTGTTGT	1850
DB	1694	CTGGAACCTCATACACTTGGTTCGAAAGTGGATTTGAACGCTGTTCTGAGTACAGC	1753
QY	1851	TCAGAGCATCTCGATTTTATTTTATAGCATTTAAAGATGCAATCTGTTGATTTGGAGAT	1910
DB	1754	TCAGAACATTTGAGATCATATTTCTCAGTTCTTAAGGGACACCATTTCTGAAACAGGAGAC	1813
QY	1911	GAAGCTTTTAAATGGCAAGCGCGATGTAACTAGCCATGTTTATTCAACTCTGTTGGAA	1970
DB	1814	AAAGCATTCACCTATCAAGGACCGCAATGTGACACACCAATTTGAAAAATTTGGTTGGAT	1873

QY	753	AGGGAATTTGGAGCAAAAAAGATGCC-----ATTCAAATGAGATCGATCGACTATTG	803
Db	312	CGGGAGATGGAATTGAAGAGCTGGCTGTGGATAGTTCTTTTGGAAAGAAAGCAATATATG	371
QY	804	GGGTATATCTCTGAGAGCACTCGGTAAATTTATATGATTTGGAATATATGGTGAAGAAATATCAG	863
Db	372	GCTTTTATCCAGAAAGGATTCGGAATAATGCTGGATCGGAATCAAGTTTATGAAGTTTCAG	431
QY	864	ATGAAAAATGGTTCCTTTTCAACTCACCATCAGCAACAGCTGCTGCTTTCATTAATCAT	923
Db	432	AGGAAGATTTGGATCATTTTTCAGCACTCCTCCACAACCTGCTGTTCGATTAATCCACAAA	491
QY	924	CAAAATCCTGGTGTGCTTAATTTAAATTCATCTTTTGGACAAGCTTTGGTAAATGCGATC	983
Db	492	TACAACGCAAGCCCTTCAATACCTAAATTTGCTTTTCAATGAATTTGGCAGTTCGAGTA	551
QY	984	CCAAACAGTTTATCTCATGATTTATTTATPCCGACTTTTCTATGGTTGACACAATTTGAAAGA	1043
Db	552	CCAGCAATGTATTCCTTCAAGGGTACATTTGTACGCTTTCATATGGTGGACGCGCTTGA	611
QY	1044	TTAGGAATTTCAACCACTTTCAGAGTGGAAATTAATAATGTTTATGATGAACAATACAGA	1103
Db	612	ATGGGAATTTCTCAGCGCTTTGTCACTGAATATGAAGCATCCCTGGCAATGCCATACAAT	671
QY	1104	TGTTGGTGGAAACGAGATGAGCAATATTCATGGATGTTTAAACATGTGCTTTAGCCTTT	1163
Db	672	TGCTGTTTACAGATGATGAGAACTCATGATGACATAGCAACATTTGCATGCGATTT	731
QY	1164	CGGTTATTAAGGATCAATGGGTATGAGTTTCCCGAGNCCOATTTGGCTGGAATTACTAAT	1223
Db	732	CGCCTTTTGGAGATGAATGGTTACGATGTTTCTCAGATGAGCTGTCTCACGTTGCTGGA	791
QY	1224	G-----AATTAGCTTTGAAAGACGAAATATGAGCTCTTGAA	1259
Db	792	GCTTCCACTTTCATGATTCATCAGAGATTTTAAATGATACAAATCCCTACTGGAA	851
QY	1260	ACATATCATGCGTCCACA---TATATTTATCAACAGAGGATTTATCTCTGGAAACAATC	1316
Db	852	TTGTACAGACCTCAAAAGTCACCTTATCAGAAAACGATCTGATCTTAGTTCGATAGGT	911
QY	1317	TTGAAGTCAGCTGATTTCTCAAGAGATATATCCAC-----	1354
Db	912	TCCTGGCTGGCAACTTATTAAGAGATAAGATGTGCTAGTAGGTGCAAAAGACTCGA	971
QY	1355	-----TGATTCAAACAGGCTTTCTTAATTAATTTACA-	1386
Db	972	TTTTTGGAGATGCTGCAACAACAAATTTAAATTTCTCATTTCCCTTGGAGGTTCAATT	1031
QY	1387	-----AAGAGGTGGAAAATGCTCTTAAGTTCCCTATCAATACCGGT	1427
Db	1032	TATGTCGTGTTTGTGNTTTCAGATCGAGTATGCTGTTAATTTCCCTTGTTATCCACA	1091
QY	1428	TTAGAACGCATAACACTAGACGAAATATACAGCTTTTCAATGTAGACAAATACAAGATT	1487
Db	1092	CTGGAGCGCTTAGAACACAAGAGAAACATCCGAACATTTTGAATGCTTGGGTTCTCTGATG	1151
QY	1488	CTGAAAACTACATATCACTCATCAAAATATTAGTAACAACATTAATACCTTAAGTTGGCTGTT	1547
Db	1152	CT---AACACAAAATCCTCATCTTTTTCGTTATCAATCAAGAAATTCCTAGCTTTGGCAGTC	1208
QY	1548	GAGATTTTACACCTGCCAATCTATTTATTCGTGAAGAAATTAAGAGTCTTCAAGAGTGG	1607
Db	1209	GAAATTTTCAGTTTCTCTCAACGTTTACCGGGATGAACTTCGGCATCTTGATAGTTGG	1268
QY	1608	GTGTTAGAGATTAAGTTGGACCGACTCAAGTTTGTCTAGGCAAAAGACCGCTTACTGTTAT	1667
Db	1269	GTGAAGAGAACAGCTGGGACGACTACAAATTTGCTCGGCAGAAAACTGACATATTCCTAT	1328
QY	1668	TTCTCTGTGTGCAACACTTTGCTCTCCCGAATATATCAGATCGGCTTATTCATGGGCC	1727
Db	1329	CTGTCTGCTGCTACCGTATTTTCTTGAAATTTGTGAOCTCGCATTTATTTATGGGCC	1388

QY	1728	AAAAATGGCATTATTAACATACAGTAGTTCGATGACCTTTTTTGGATATCGGTGGTACAAATCGAT	1787
Db	1389	AAAAATGGGTGTCCTCACAACTGTGGTTGATGACTCTTCGGATGTGGTGGATCAAAAGAA	1448
QY	1788	GAAATTGACCAACTGATTCATGTGTTGCAAAAAAGGAATGTAGATGTGCGAACAGGATGT	1847
Db	1449	GAAATTAGAAAACTCATAGCTACTAGTTGAGAAATGGCATGGCCACCATGCAGTTGAGTTC	1508
QY	1848	TCTTCAGAGCATGTTTCGGATTTTATTTTTTTAGCATATAAAGANGCAATCTGTTCGATTGGA	1907
Db	1509	TATTCGGAACAGGTGAATAATAGTATTTTCTGCTATTTATACACAGTGAACCACTCTTGGA	1568
QY	1908	GATGAAGCTTTTAAATGCGAAGCGCGGATGTAACCTAGCCATGTTATTCAAAACCTTGTTG	1967
Db	1569	GCAATGGCTTCTGCAGCACAAAGCGCTGATCTTACAAACCACTAGTAGAAAATATGCGCTG	1628
QY	1968	GAACTAAATGAATAGTATGCTTCGAGAGAGCTATATGCGAACAGAGATGCTTATGTGCCAACA	2027
Db	1629	GAATTTGTTAAGATCTATGATGGTCTGAGGCGAATGGCGAGAGATGCCAATAATGTACCAACA	1688
QY	2028	TTAAATGAATATATGGAAGACGCTTACGTTGTCATTTGCAATTTAGGCCCGATTTGTCAAGCCG	2087
Db	1689	GTTGAAGAATACATGACAAATGCTGTGTTGCTCATTTGCACTGGGCCCAATGTGCTCCCA	1748
QY	2088	GCTATTACTTGTGGGGCCCAATATATCAGAGAGATGTTGAAAGCTCTGTAATATCAT	2147
Db	1749	GCAATGTATTTGTAGGGCGAAGCTATTAGAGCATGCTGTCAAAGATGAAGAGTAGCAT	1808
QY	2148	AATCTATTTAAGCTAANTAGCACGACGCGGTGCGACTTCTAAACGATATCCATAGCTTCAAG	2207
Db	1809	AAATTAATTTAGGCTAGTAGGACACTTTCGGGAGGCTCCTCAATGACTACCAAAAGTTTAGAG	1868
QY	2208	AGGGAATTTTAGGAAGCGCAATTAACCGCGTAGCATTTGCATTTGATGTAACGCGAAGCT	2267
Db	1869	AGSGAAGGCAACACAGGGGAAGCTCAATAGTGTTCCTACTTGTGCTCCACAGTGTGTT	1928
QY	2268	GGGAAAGTGGGAAGAAGAGTTGTGGAGGAGATGATGATGATGATTAATAAACAGAGGAAA	2327
Db	1929	TCTATGTCCATAGAAGCGCGCTAAAAAGGCAATGCAGAAGTCCATAGACGTGCTAGGAGA	1988
QY	2328	GAATTAATGAATTAATTTTGAAGAAATGGTAGCATTTGTTCTCTAGAGCTTTGTAAGAT	2387
Db	1989	GACTTGCTATAGATGGTTCT--CAGGAAAGAAAGTGTGTTCTCTAGGCCATGCAAGGAG	2045
QY	2388	GCATTTTGGAAACATGTCTACAGTGTGAATTTTTTTTACGCAAAACGATGACGGGTTTA	2445
Db	2046	CTCTCTCGAAGATGTGAAGATCTTCACTGTTTACTTCAGAAATGATGATTTA	2103
RESULT 10			
ADA70130			
ID	ADA70130 standard; DNA; 2154 BP.		
AC	ADA70130;		
XX			
DT	20-NOV-2003 (first entry)		
XX			
DE	Rice gene, SEQ ID 3453.		
XX			
KW	Plant; bacterial infection; fungal infection; viral infection; rice;		
KW	gene; ds.		
XX			
OS	Oryza sativa.		
XX			
PN	W02003000898-A1.		
XX			
PD	03-JAN-2003.		
XX			
PF	22-JUN-2001; 2001WO-IB001105.		
XX			
PR	22-JUN-2001; 2001WO-IB001105.		
XX			
PA	(SYGN) SYNGENTA PARTICIPATIONS AG.		

XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
DR
XX Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 6; SEQ ID NO 3453; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC the expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interraction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2154 BP; 596 A; 465 C; 556 G; 537 T; 0 U; 0 Other;
Query Match 14.9%; Score 416.8; DB 8; Length 2154;
Best Local Similarity 52.7%; Pred. No. 4.9e-88;
Matches 1109; Conservative 0; Mismatches 937; Indels 60; Gaps 7;
QY 381 ATGGTCCCTCTCCAACTCAACCAATCCCTGTTTCCCTGAGTGTCTCAATTGGTTA 440
DB 1 ATGGTCCCACTCCGGGGTTCGTCTCAATCCAACTCCAAAGCTTCCCTCAATGCGTGGATG 60
QY 441 ATTAATAATCAGCTTAATGATGGTTTCATGGGGTCTTGTAAATCACACTCATATATCAAT 500
DB 61 CTAGAGATCAATGGGACATGATGATCATGAGTATTGACGGGTCCATATCCACAGCCAA- 119
QY 501 CACCCTGTCTTAAAGATCTCTATCTTCAACATAGCATGATTTGTTGATTAATAAGA 560
DB 120 -----CAAAGATGTCCTATCATCTACGCTAGCATGTGTTGCGCACTCAACAAA 168
QY 561 TGGAAATGTTGGGAGATCAATAATAAAGTCTCAAGTTTATTAGTCAAAATCTTGCT 620
DB 169 TGGAAATGTTGGTGGGAGCATCATCAGAGAGGACTAAGTTTCAITGGAAGAAATTTCTG 228
QY 621 TCAGCTACTGAAAAAAGTCAACCACTCCCAATGTTGTTTGAATCATATTTCTCTGTTG 680
DB 229 ATTGGATGATGATCAAGCAGTTGCTCTATAGTTTTCGGCATCACTTTCCAGCTATG 288
QY 681 CTTGAGTATGGAATAAATTGGACATTAACCTCTTTCAAAACAAACAGATTTTATGTTG 740
DB 289 CTAACCTTGGCAATGGTCCGGTCTGGAAGTCCCGTTCAGACAAAACGATATTGATAGT 348
QY 741 ATGCTATAGAGGGAATTCGACAAATAAGA-----TGCCATTCAAATGAGATG 791
DB 349 CTTAATCACCTTCGGGAGATGAATAATCAAGAGAGCGCGGAACCAATCTCTGTTGAAGA 408
QY 792 GATGGATCTTGGGCTATATCTCTGAAGGACTCGGTAAATTTATATGATTTGGAATATGGT 851
DB 409 AAGGCTATATGGCTATCTCGCAGAGAGATTTGGGAACCTTCTGGAATGGGATGAATC 468
QY 852 AAGAAATATCAGATGAAAGAGTTCGTTTTCACACTCAACCATCAGCAACAGCTGCTGCT 911
DB 469 ATGATGTTCCAGAGGAAAGCGGTCATTGTTCAACTGTCTTCTTCAACTGCTGGCGCA 528
QY 912 TTCATTATATCAAAATCTCGTGTCTTAAATTTAAATTTCACTTTTGGACAAGTTT 971
DB 529 TTAGCCCAATACCAGCAGATTAAGCTCTCCAACTATTGCAATCTCTAGTCAATAAATTT 588
QY 972 GGTAAATGAGTCCCAACAGTTTATCCTCATGATTTATTTATCCGACTTTTATGTTGAC 1031
DB 589 GACGCTGTAGTGCCAAACACTGTATCCACTAAACATATATTGTTCAGCTTTCAATGGTGGAT 648

QY 1032 ACAATTGAAAGATTAGGAATTTCAACCAATTTACAGATGGAATTTAAATAATGTTTTAGAT 1091
DB 649 GCGCTTGAACATCGGAATATCTCAGTATTTTGCAGTGAAATAAGAGCATCTCTGAC 708
QY 1092 GAAACATACAGATTTGGTGGAAACGAGATGAGCAAAATTAATCATGATGTTGTTAATG 1151
DB 709 ATGACATACAGTTCTCTGTTAGGAAAGATGAGGAATCATGTTAGACGTAACACGTC 768
QY 1152 GCTTTAGCTTTCCGTTTATAGGATCAATGGTATGAAGTTTCCACATCATCATGGCT 1211
DB 769 GCAATGGCAATTTCCGCTTTTAAAGAAATGAACGGTTACGATGTTTCTCAGATGACTATCT 828
QY 1212 GAAATTAATAAATAAGTTAG-----CTTTGAAAGACGAAATAT 1247
DB 829 CATGTTGCGGAGCTTCAGGCTTCGTTGATTCGTTCAAGGTTATCTAAACGATAGAAA 888
QY 1248 GCAGCTCTTTGAAACATATCATCGCTC---ACATATATTAACAAGAGATTTATCTTCT 1304
DB 889 TCTGTACTGGAAGTATACAAAGACCTCGAAACACAGTATATCAGAAATGATTTGATCTTA 948
QY 1305 GGAACAAATCTTGAAGTCAGTATTTCTCAAGAGATTAATATCCACTGATTTCAAAAC 1364
DB 949 GATAGCATAGGCTCTTGGTCAGGACCTATTCAAGAAATGCTGTGCTTAATGGGAAA 1008
QY 1365 AGGCTTTCTAAATTAATTCACAAAGAGGTGGAATGCTTTAAGTTCCCTATCAATACC 1424
DB 1009 GGAATCTC-----AGGCGGTGAAGAGATCGAGTTTGTCTGAAGTATCCCTTCTACAGC 1062
QY 1425 GGTTTAGACGCAATAAACACTAGACGAATATACACTTTCAATGATAGACATAAAGA 1484
DB 1063 ACATTGGAACGCTAGTCCACAGAAAGAACTTTGCTTTTCGACGCTAGGCTCCAG 1122
QY 1485 ATTCTGAAACTACATATCACTCATCAATATATTAGTAAACACTGATTAACCTTAAGTTGGCT 1544
DB 1123 ATGCTGAAACAGAGTGCATGCTGTTTATGACAGTCA---AGATTTCTAGCTTTGGCT 1179
QY 1545 GTTGAAGATTCTTACACCTGCCAATCTATTATTCGTGAAGAAATTAAGGCTCTTGAAGG 1604
DB 1180 GTGATGATTTCTGCACTCTCAATCTAATTAATCCAGAAATGAACTGAATTTATCTGAAAGT 1239
QY 1605 TGGGTGATGAGAAATTAAGTTGGACCACTCAAGTTTGTAGGCAAAAGACGCTACTCT 1664
DB 1240 TGGGTGAAAGACACAGACTCGACCACTCATTTGCGACGGCAGAAAGATAACATATGC 1299
QY 1665 TATTTCTCTGCTGCTCAACACTTTGCTCTCCGAAATTAATCAGATGCGCGCTATTTCAATGG 1724
DB 1300 TATCTCTCTGTTGCTGCCACCACTTCCTGCTCTGAAATGGGCTACGCTCGACCTCTGG 1359
QY 1725 GCCAAAAATGGCATATTAACTACAGTATGTTGATGACTTTTGTGATATCGGTGGTACATC 1784
DB 1360 GCAAGAACCGCTTGGTTGACGGCTGTTATCGAAGATCTCTTCGATGTTGGTGAATTAGAA 1419
QY 1785 GATGAATGACCAACCTGATTCATGTTGTTGAAATGGAATGATGATGTCGACAAAGAT 1844
DB 1420 CAAGAAACAAGAAACCTTCTAGCATTAATGGAGAAGTGGGAAGAGCCTGTTGAAGATGAG 1479
QY 1845 TGTGTTTCAGAGCATGTTCCGATTTTATTTTAGCATTTAAAGATGCAATCTGTTGGATT 1904
DB 1480 TACTACTCTGAGATGTAAGATCGTGTTCAGGCTCTTTTATAATAACGTTGAACGAGATT 1539
QY 1905 GGAGATGAAGCTTTTAAATGCAAGCGCGGATGTAATCTAGCCATGTTTATTTCAAACTGG 1964
DB 1540 GGCAGAAAGGCTTCGCGGTTTACAGGCGCATGATGTTTACCAAGTACCTTAGTAGACGTATGG 1599
QY 1965 TTGAACTAATAGATGATGTTGAGAGAGACTATATGGAACAGAGATGCTTATGTGCCA 2024
DB 1600 CTACATGTTGGTGGTGTATGAAGTTGAGGCAAAATGGCAGAGAGCCAGCACTGCCA 1659
QY 2025 ACATTAAATGAATATATGGAACCGCTTACGTGTCAATTTGATTTAGGCCCGGATTTGTCAG 2084
DB 1660 ACGTTGGAAGATACATGAAAGTGGCATGTTGTTCACTTGGACAGGGCTGTACCGGTGATG 1719

QY 2085 CCGGCTATTACTTTGCGGCCCAAAATATATCAGAGAGATGTTGAAAGCTCTGAATAT 2144
Dd |||||
1720 TCGGCAATGTTCTGATCGGAGAAAAGCTCCGGAGGGTATTCTCGAATTGAGATATC 1779
QY 2145 CATAATCTATTAAAGTAATGACGACGAGGTCGACTTCTTAAACGATATCCATAGCTTC 2204
Dd |||||
1780 GATGAGTTGTTTCAAGGCTGATGGGCACTTGGCGCGTCTCTCGAATGACATTCGAGGCATT 1839
QY 2205 AAGAGGGAATTTAAGGAAGCAAAATTAACGCGGTAGCATTTGATTTGAGTAACGGAGAA 2264
Dd |||||
1840 GAGAGGAGAGTCGAGCGGCAAAATGACGACGGGCTCTCGTCTCGTTACCGTAGC 1899
QY 2265 AGTGGAAAGTG--GAAGAAGAGGTTGTGAGAGATGATGATGATTTAAACAAAG 2321
Dd |||||
1900 GCGGCTCCATGTCCTAGACGAGGCCAAAACCGAAGTGTAGGCGCATCGAGCCTCA 1959
QY 2322 AGGAAGAATTAATGAAATTAATTTTGAAGAAATGATGATGATTTCTTCTAGAGTTGT 2381
Dd |||||
1960 CGAGGAAAGCTGCTGAGCTTGGTCTGTCAGCGAAACAGGAAGGCCCTATCCGAGCGCTGC 2019
QY 2382 AAAGATGCAATTTTGAACATGTGTACGNGTTGAATTTTATACCAACGATGACGG 2441
Dd |||||
2020 AGCAGCTGTTCTGGAAGATGTGCAAGATCTTCACCTGTTCTACTACCAGACCGGACGA 2079
QY 2442 TTACT 2447
Dd |||||
2080 TTCACT 2085

RESULT 11

AD69971

ID AD69971 standard; DNA; 2223 BP.

XX

AC AD69971;

XX 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 3294.

DE Plant; bacterial infection; fungal infection; viral infection; rice;

KW gene; ds.

XX Oryza sativa.

XX WO2003000898-A1.

XX 03-JAN-2003.

XX 22-JUN-2001; 2001WO-IB001105.

XX 22-JUN-2001; 2001WO-IB001105.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

XX Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX WPI; 2003-175290/17.

XX Identifying at least one gene involved in plant resistance or response to

XX pathogenic infection for conferring resistance or tolerance to a plant to

XX bacterial, fungal or viral infection by determining or detecting plant

XX gene expression.

XX Claim 6; SEQ ID NO 3294; 899pp; English.

XX The present invention relates to a method (M1) for identifying genes

XX involved in plant resistance or response to pathogenic infection. M1

XX comprises identifying a gene whose expression is significantly altered in

XX the incompatible interaction of plant gene expression relative to

XX expression of the gene in an uninfected plant, in a mutant plant that

XX does not express a gene associated with response to pathogenic infection,

XX or in a corresponding incompatible or compatible interaction. (M1) is

CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.

XX

SQ Sequence 2223 BP; 552 A; 528 C; 634 G; 509 T; 0 U; 0 Other;

Query Match 13.6%; Score 379.4; DB 8; Length 2223;

Best Local Similarity 54.5%; Pred. No. 3.4e-79;

Matches 900; Conservative 0; Mismatches 676; Indels 75; Gaps 4;

QY 795 GGATATCTGGCGTATATCTCTGAAGACTCGGTAAATTTATATGATTGGAATATGTTGAAG 854

Dd |||||

574 GAATATACGGCTTATGCTCTCAGAGGATTAGCAATATTCAGACTGGATGAAGCATG 633

QY 855 AAATATCAGATGAATAATGGTTCTGTTTCACTCACTACCATCAGCAACAGCTGCTGTTTC 914

Dd |||||

634 AAATTTCCAAAGAAAGAAATGGCTCACTGTTTCAACTCCCTTACACAACTGCAGTGCATTA 693

QY 915 ATATATCATCAAAATCCCTGGTTGTTCTTATATTTAAATTTCACTTTTGGACAGATTGGT 974

Dd |||||

694 GTCCCAACTATGATGCCAAAGCTCTCCAGTACTTAGACATGCTTCTGGACAAATTTGA 753

QY 975 AATGCACTCCCAACAGTTTATCTCATGATTTATTTATCCGACTTTCTATGTTTGCACCA 1034

Dd |||||

754 AGTGCAGTGCAGCGGCTATCTGCAATATTCAGTCTCAGCTCTACATGGTGGATG 813

QY 1035 ATTGAAGATTTAGGAATTTTACACCATTTTCAGAGTGGAAATTTAAATATTTTGTAGTAA 1094

Dd |||||

814 CTTGAAAAGATGGGAATATCTAGGCATTTTGTGGTGGATAAAGAGCATACTGGACATG 873

QY 1095 ACATACAGATGTTGGTGGAAACGAGATGAGCAAAATTTATTCATGATGTTTGTACATGCT 1154

Dd |||||

874 ACCTACAGTTGCTGGAACAGAGGATGAGGAAATTTGCTTGACATGCAAAACATGTTGG 933

QY 1155 TTAGCCTTTGCGTTTATTAAGGATCAATGGGTATGAAGTTTCCCGAGATCCATGGCTGAA 1214

Dd |||||

934 ATGGCATTTGCTGATGTTGGTATGATGATGATGATGATGATGATGATGATGATGATGAT 993

QY 1215 ATTACTAATGAATTTAGCTTTTGAAGACGAATATGAGCTCTTTGAAACATATCATGCTCA 1274

Dd |||||

994 TTTTCT-----GAACTTTCAAGTTTCCCACAATTTCA 1023

QY 1275 CATATATTATCCAAAGAGATTTTATCTTCTGGAACCAAAATCTTGAAGTCAGCTGATTTTC 1334

Dd |||||

1024 CTTCAAGGATATCTGAATGATACAAAGATCTTTATTAGAATTTACATAAGCTTTCAAAAGTC 1083

QY 1335 CTCAAAGAGATAATATCCACTGATTTCAAAACAGGCTTTCTAAATTAATTTTCAAAAGAGGTG 1394

Dd |||||

1084 -----AGTATCGGAAAGAGGTTT 1104

QY 1395 GAAATGCTCTTAAAGTTCCTTATCAATACCGGTTTGAACCGAATAAACACTAGACGAAT 1454

Dd |||||

1105 GAATATGCTCTTGAATTTCCCTTCTATACCATCTTGGATCTGTAGACCATAAAGAAAT 1164

QY 1455 ATACAGCTTTCAATATGATAGCAATACAGAAATTTCTGAAACTACATATCACTCATCAAT 1514

Dd |||||

1165 ATCGAACATTTTGACATTTACAGAGTCAAGTGTAGAAACAGGCT---ACTTGCATGT 1221

QY 1515 ATTAGTAACACTGATTTACTTAAAGTTGGCTGTTGAAAGATTTTCTACCTGCGCAATCTATT 1574

Dd |||||

1222 CATTCCAATGAAGAAATCATGCGCTTGGGTGTGAGAGATTTTAGTACTCTCAGTTTATT 1281

QY 1575 TATGCTGAAGATTTAAAGGTTTGAAGGTTGGGTGTAGAGATAATAGTTGGACAGCTC 1634

Dd |||||

1282 TTCCAAGAGAGCTGCAGCAACTCAACAGCTGGGTGAAAGAGAGCAGGTTCGATCAGCTG 1341

QY 1635 AAGTTTGTAGCAAAAGACCGCTACTGTTATTTCTCTGTGTGCAACACTTTCGCTT 1694

Dd |||||

1342 CAATTCGACCGGCAAGATTTGGAATCTTCTATTTCTCTGCTGTGCTGCTACCAATTTTCACT 1401

QY 1695 CCCGAAATTTATCAGATGCGGCTATTTTCATGGGCCAAAATGCAATATTAACCTACAGTAGTT 1754

Dd |||||

1402 CCTGAACCTGTCAGATGTTGCGATTTTGTGGCCAAAATGCGCTGCTGCAACCGTCTGCTC 1461

QY 1755 GATGACTTTTGTATCGGTGCTAATCGATGATGACCACTGATTCATGCTGT 1814
 Db 1462 GAGCACTTCTTCGCGTGGAGATCAAAAGAAGAACTGGAAAACCTCTGCGCATAGTT 1521
 QY 1815 GAAAATGAATGTAGATGCTGCAAGAGATTTGTTTCAGAGCATGTTGCGATTTTATTT 1874
 Db 1522 GAGAGTGGGCAAGATGACAAACTGAGTACTCTCTGACAACTAGAGATGTTGTTT 1581
 QY 1875 TTAGCATTTAAAGATGCAATCTGTTGGATGAGATGAAAGCTTTTAAATGGCGAGCGCG 1934
 Db 1582 TCTGCAATTTATACCTTCAACTAACCACTTGGATCAATGGCTCTCTAGTACAAAGCCGT 1641
 QY 1935 GATGTAACCTAGCATGTTTATCAAACTGTTGGTGGAACTAATGAATAGTATGTTGACAGAA 1994
 Db 1642 GATGTCACCAACACCTTTAGAAATATGGCAAGAAATGCTGAGGCTTATGATGACAGAG 1701
 QY 1995 GCTATATGCAACAGAGATGTTTATGTCACCAACATTAATGAATATATGAAAACGTTTAC 2054
 Db 1702 GTAGAGTGGAGGAGAGCGGTATGTCACACAGCAGAGGAATACATGGAATGCAATGTT 1761
 QY 2055 GTGCTATTTGCAATGAGCGGATGTCACCGCGCTATTTACTTTGTTGGGGCCCAATTA 2114
 Db 1762 GTGACATTTGCACTGGGACCGGTTGTCTCCAGCATTTGTATCTTGTGGACCAAGATC 1821
 QY 2115 TCAGAGGAGATTTGAAAGCTCTGAATATCATATTAATTAAGCTAATGACACGCGAG 2174
 Db 1822 CCGTATCTGTCTAAGAACCAAGAAATGCGAGGTTGTTCCGGCTAATGACAAATGT 1881
 QY 2175 GGTGCGACTTCTAAACGATATCCATAGCTTCAAGAGGGAATTTAAGGAAGCAAAATTAAC 2234
 Db 1882 GCGCGTCTCTGATGATGTCATCTACGAGAGAGGCGCAGCCAGGSCAAGCTGAAC 1941
 QY 2235 GCGGTAGCATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2294
 Db 1942 AGCGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2001
 QY 2295 GAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2354
 Db 2002 CAGATTCAGAGACCCATCGAATGATGATGATGATGATGATGATGATGATGATGATGAT 2058
 QY 2355 AATGTTAGCATGTTCTCTAGAGCTTTGAAAGATGCAATTTTGGAAATGTTGCTGCTGTTG 2414
 Db 2059 AGAGGAGGCGCGTTCACAGGCCATGAGGAGCTGTTCTGGAGCATGTCGAAGTCTGC 2118
 QY 2415 AATTTTTCAGCAACGATGACCGGTTTA 2445
 Db 2119 CACTTCTTCTACTCCGCGCGGCGGTTTCA 2149

RESULT 12
 ADI45176
 ID ADI45176 standard; cDNA; 2223 BP.

XX AC ADI45176;

XX DT 22-APR-2004 (first entry)

XX DE Rice isoprenoid biosynthesis-associated cDNA #54.

XX KW Rice; isoprenoid biosynthesis; ss; gene; plant; isopentenyl diphosphate;
 KW IPP; dimethylallyl alcohol; DMAPP; short-chain plastid prenyltransferase;
 KW gibberellin; carotenoid; abscisic acid; tocopherol; plastoquinone;
 KW phyloquinone; mevalonate pathway; phytosterol; brassinosteroid;
 KW haeme; yield.

XX OS Oryza sativa.

XX PN US2004010815-A1.

XX XX 15-JAN-2004.

XX

PF 26-SEP-2002; 2002US-00259194.
 XX 26-SEP-2001; 2001US-0325277P.
 PR 04-APR-2002; 2002US-0370620P.
 PR 04-APR-2002; 2002US-0370743P.
 XX (LANG/) LANGE B M.
 PA (GHAS/) GHASSEMIAN M.
 PA (BRIG/) BRIGGS S P.
 PA (COOP/) COOPER B.
 PA (GLAZ/) GLAZEBROOK J.
 PA (GOFF/) GOFF S A.
 PA (KATA/) KATAGIRI F.
 PA (KREP/) KREPS J.
 PA (MOUG/) MOUGHAMER T.
 PA (PROV/) PROVART N.
 PA (RICK/) RICK D.
 PA (ZHUT/) ZHU T.
 XX Lange BM, Ghassemian M, Briggs SP, Cooper B, Glazebrook J;
 PI Goff SA, Katagiri F, Kreps J, Moughamer T, Provart N, Ricke D;
 PI Zhu T;
 XX WPI; 2004-090562/09.
 DR P-PSDB; ADI45177.
 XX New isolated polynucleotides and polypeptides associated with isoprenoid
 PT synthesis in plants, useful for producing transgenic plants, for targeted
 PT gene disruption, as well as markers or probes.
 XX Claim 1; SEQ ID NO 107; 117pp; English.
 PS The invention relates to a polynucleotide (or its complement, protein
 CC encoding fragment or reverse complement), comprising a nucleotide
 CC sequence encoding a polypeptide comprising an amino acid sequence
 CC involved in or associated with the biosynthesis of isoprenoids in a rice
 CC plant. Also included are an isolated polypeptide involved in or
 CC associated with the biosynthesis of isoprenoids in a plant, an expression
 CC cassette comprising the polynucleotide, a host cell comprising the
 CC expression cassette, and a transgenic plant comprising the expression
 CC cassette. The polypeptides and polynucleotides include those associated
 CC with the biosynthesis of isopentenyl diphosphate (IPP) and dimethylallyl
 CC alcohol (DMAPP), the biosynthesis of short-chain plastid
 CC prenyltransferases, the biosynthesis of gibberellins, the biosynthesis of
 CC carotenoids and/or abscisic acids, the biosynthesis of tocopherols,
 CC plastoquinone and/or phyloquinone biosynthesis, the mevalonate pathway,
 CC phytosterol and brassinosteroid metabolism, biosynthesis of ubiquinone,
 CC biosynthesis of chlorophyll or haeme. Also disclosed are banana, wheat
 CC and corn homologues of some of the rice polynucleotides. The
 CC polynucleotides are useful for producing transgenic plants, where the
 CC genome is augmented by a nucleic acid molecule of the invention, or in
 CC which the corresponding gene has been disrupted, e.g. to result in a
 CC loss, a decrease or an alteration in the function of the product encoded
 CC by the gene. The plants may also have increased yields and/or produce a
 CC better quality product than the corresponding wild-type plant. The
 CC nucleic acid molecules are useful for targeted gene disruption, as well
 CC as markers and probes. Note: The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?docID=20040010815. The present sequence
 CC is a Rice isoprenoid biosynthesis- associated cDNA of the invention.
 XX Sequence 2223 BP; 552 A; 528 C; 634 G; 509 T; 0 U; 0 Other;

Query Match 13.6%; Score 379.4; DB 12; Length 2223;
 Best Local Similarity 54.5%; Pred. No. 3.4e-79;
 Matches 900; Conservative 0; Mismatches 676; Indels 75; Gaps 4;

QY 795 GGATACCTTGGCTATATCTCTGAAGGAGCTCGGTAATTTATATGATTCGAATATGCTGAAG 854
 Db 574 GAATATACGGCTTATGTTCTCAGAGGATTAGGCAATATTCAGAACTGCAATGAATGATG 633

1270	CGTCACATATATATACCAAGAGGATTTATCTTCTGGAAAAACAATCTTGAAGTCACGTCG	1320
116	CGACAGTTAGTATCTCTGAAGATGAGTCTATCCTGGATAGCATAGGCTCAAGGTCACGTA	175
1330	ATTTCCTCAAAAGAGATAATATCCATCGATTCAAAACAGGCTTTCTAAATTAATTCACAAG	1389
176	CCTTACTGAGGGAACAACCTAGAGTCTGGTGGTGCTCTACGAAAACCTTCACTCTTTAAAG	235
1390	AGGTGAAAAATGCTCTTAAAGTTCCCTATCAATACCGGTTTAGAAGCATAAAGCACTAGAC	1449
236	AGGTGAACATGCTCTGGAAGCGTCCCTCTACACACATTTGGACCGCTCTACACCATAGGT	295
1450	GAATATACAGCTTTTCAATGTAGACAATACAAAGAAATCTGAAAACTACATATCACTCAT	1509
296	GGAACTCGAAAATTTCAATATTTATAGACGACCATGCTAGAGACACCATACTTGTCAA	355
1510	CAAAATATTAGTAAACATCATGATTAGTTCAGTTGGCTGTTGAAGATTTCTACACCTGCCAAT	1569
356	ATCAACATATCCAGTAGAGATATTTCTAGCGTTGAGATTTAGAGACTTCAGTTCCCTCAGT	415
1570	CTATTTATTCGTGAAGAAATTTAAAAGGTCCTTGAAGGTGGTGCTAGAGAATAAGTTGGACC	1629
416	TTACTTACAGCAAGAACTTCAACATCTTGAAGCTGGTGAAGAGTGCAAGTTAGACC	475
1630	AGTCAAGTTTGTAGGAAAAAGACCGCCCTACTGTATTTTCTCTGCTGCTGCAACACTTT	1689
476	AGCTACAAATTTGGCGACACAGAGTTGGCATACTTCTACTTGTCTGCTCGCCCAATGT	535
1690	CGTCTCCGAATTTATCAGATGCGGPAATTTCAATGGGCCAAAAATGGCATATTAATCAAG	1749
536	TCCTCTCTGAGCTGCTGATGTCGGAACCTTTGTGGCCAAAAATGTTGTGCTCAACAATA	595
1750	TAGTTGATGACTTTTTTGTATATCGGTGGTACAAATCGATGAATGTGCCAACTGATTCAAT	1809
596	TTGTTGACCACTCTTTTGATGTTGCCGGATCAAAAAGAAACCTTCAAAAACCTTTGTCATGT	655
1810	GTCTTGA AAAATCGAATGTAGATGTCGACAAAGATTTGTTTTCAGAGCATGTTGCGATTT	1869
656	TGGTTGAGATGTCGGGACGACATCAAAAGTTGAGTTCTACTCAGAAACAAGTAGAGATTA	715
1870	TATTTTTAGCAATTTAAAGATGCAATCTGTTGGATTTGGAGATGAAGCTTTTAAATGGCAAG	1929

OS	Oryza sativa.
XX	
PN	WO2003000898-A1.
XX	
PD	03-JAN-2003.
XX	
XX	22-JUN-2001; 2001WO-IB001105.
PF	22-JUN-2001; 2001WO-IB001105.
XX	(SYGN) SYNGENTA PARTICIPATIONS AG.
XX	
PA	Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI	Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
PI	WPI; 2003-175290/17.
XX	
PT	Identifying at least one gene involved in plant resistance or response to
PT	pathogenic infection for conferring resistance or tolerance to a plant to
PT	bacterial, fungal or viral infection by determining or detecting plant
PT	gene expression.
XX	
PS	Claim 6; SEQ ID NO 3295; 899bp; English.
XX	
CC	The present invention relates to a method (M1) for identifying genes
CC	involved in plant resistance or response to pathogenic infection. M1
CC	comprises identifying a gene whose expression is significantly altered in
CC	the incompatible interaction of plant gene expression relative to
CC	expression of the gene in an uninfected plant, in a mutant plant that
CC	does not express a gene associated with response to pathogenic infection,
CC	or in a corresponding incompatible or compatible interaction. (M1) is
CC	useful for conferring resistance to resistance or tolerance to a plant to
CC	bacterial, fungal or viral infection. The present sequence was used to
CC	illustrate the invention.
XX	
SQ	Sequence 2193 BP; 593 A; 440 C; 564 G; 596 T; 0 U; 0 Other;
Query Match 11.3%; Score 315.6; DB 8; Length 2193;	
Best Local Similarity 58.1%; Pred. No. 3.9e-64;	
Matches 615; Conservative 0; Mismatches 434; Indels 9; Gaps 3;	
QY	1388 AGAGGTGGAAAACTCTTAAGTTCCCTATCAATACCGGTTTAGAAGCAGATAAACACTAG 1447
Db	1044 AAAAATGAAGTATGCTCTTAAGTTTCCTTCTCACCAACACTGACCCTGAGTACAA 1103
QY	1448 ACGAAATATACAGCTTTTACAATGAGACAAATACAGAATTCTGAAAACATCATATCACTC 1507
Db	1104 GAGGAACATTGAACGTTTTTGATGCAAAGGATTCACAGATGTTAAAGACGGAAT--ACTT 1160
QY	1508 ATCAAATATTAGTAACACTGATTAACCTTAAGTTGGCTGTTGAAAGATTTTACACCTGCCA 1567
Db	1161 GCFTTCCTCATGCCAATCAAGATATTTAGCTTTGGCTGTTGAAGATTTTCAGTAGTTCTCA 1220
QY	1568 ATCTATTTATCGTCAAGAAATTAAGAGTCTTGAAGGTGGGTGGTGAAGATAAGTTTGGGA 1627
Db	1221 ATCTATATACAGATGAACTTAATTATCTTCACTGTTGGTGAAGAGATGAAAGCTCGA 1280
QY	1628 CCAGCTCAAGTTTCTAGGCAAAAGACCGCCTACTGTATTTCCTGTCTGTCTGCAACACT 1687
Db	1281 TCAGCTGCCAATTTGCAAGCCCAAGAGTTGACATAATGCTTACCTTTCTGCTGCTAGCAT 1340
QY	1688 TTCGTCTCCGAATTATCAGATCGCGTATTTTCATGGGCCAAAATGCGCATATTAACTAC 1747
Db	1341 ATTCCCCCGTGAATTTGCTGAAGCCCAGATTGCAATGGCTTAAATGTTGTTACTGACAC 1400
QY	1748 AGTAGTTGATCACTTTTTTTTGATATCGGTGGTACCAATCGATGAATTGCCAACCTGANTCA 1807
Db	1401 TGTTGTTGATGACTCTTTTGATCTTGCGGGATCAAAAGAGAACTAGAAAACCTTCATTGC 1460
QY	1808 ATGTGTTGAAAAATGGAAATGATGATGTCGACAAGATTTGTTGTTTCAGAGCATGTTCCGAT 1867
Db	1461 TTTAGTTTCAGAGATGGGATGACAT---CAAGAGAGATTTCTACTCAGAACAAAGTAAGAAT 1517

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 15:57:41 ; Search time 11750.6 Seconds
(without alignments)
11236.263 Million cell updates/sec

Title: US-10-041-018-361
Perfect score: 2792
Sequence: 1 cccactcatcctttatcaac.....aaaaaaaaaaaaaaaaaaaaa 2792

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2792	100.0	2792	8 AF097311	AF097311 Stevia re
2	2582.6	92.5	3117	8 AF097310	AF097310 Stevia re
3	1428.6	51.2	2594	8 AB031205	AB031205 Lactuca s
4	809.4	29.0	2450	8 AB045310	AB045310 Cucumis s
5	782.2	28.0	2858	6 E12936	E12936 cDNA encodi
6	782.2	28.0	2858	8 CM043904	CM043904 Cucurbita m
7	732.4	26.2	2506	8 AF034774	AF034774 Arabidops
8	559.8	20.1	2271	8 AY347876	AY347876 Oryza sat
9	555	19.9	2283	8 AY347877	AY347877 Oryza sat
10	543.4	19.5	2283	8 AY347878	AY347878 Oryza sat
11	530.8	19.0	3234	8 AK119442	AK119442 Oryza sat
12	479.6	17.2	2529	8 AY347880	AY347880 Oryza sat
13	467.6	16.7	2753	8 AK119327	AK119327 Oryza sat
14	453.2	16.2	2403	6 AR452826	AR452826 Sequence
15	448.2	16.1	3061	8 AK068310	AK068310 Oryza sat
16	444.6	15.9	2772	8 AB089272	AB089272 Oryza sat
17	434.8	15.6	2497	8 AY347882	AY347882 Oryza sat
18	416.8	14.9	2154	6 AX653583	AX653583 Sequence
19	411.4	14.7	2888	8 AK121446	AK121446 Oryza sat

20	386.4	13.8	2813	8 AK072461	AK072461 Oryza sat
21	379.4	13.6	2223	6 AX653424	AX653424 Sequence
22	378	13.5	2220	8 AY347879	AY347879 Oryza sat
23	368.6	13.2	2463	8 AB118056	AB118056 Oryza sat
24	355	12.7	1773	8 AF529266	AF529266 Zea mays
25	332.4	11.9	2149	8 AY347881	AY347881 Oryza sat
26	330	11.8	1559	8 AF105149	AF105149 Zea mays
27	329.4	11.8	1875	6 AX653521	AX653521 Sequence
28	315.6	11.3	2193	6 AX653425	AX653425 Sequence
29	241.6	8.7	2313	6 AX653751	AX653751 Sequence
30	238.2	8.5	1157	6 AR452831	AR452831 Sequence
31	194.8	7.0	2861	6 BD227676	BD227676 Synthesas
32	194.8	7.0	2861	6 AR240711	AR240711 Sequence
33	194.8	7.0	2861	6 AR266986	AR266986 Sequence
34	194.8	7.0	2861	6 AR316335	AR316335 Sequence
35	194.8	7.0	2861	6 AR338478	AR338478 Sequence
36	194.8	7.0	2861	6 AR429884	AR429884 Sequence
37	194.8	7.0	2861	8 AGU50768	AGU50768 Abies grand
38	166.6	6.0	2643	8 AY424738	AY424738 Taxus bac
39	166.6	6.0	2771	8 AY461450	AY461450 Taxus x m
40	160.2	5.7	1568	8 AK108710	AK108710 Oryza sat
41	158.6	5.7	2706	8 AY007207	AY007207 Taxus chi
42	153.8	5.5	2700	6 AR110171	AR110171 Sequence
43	153.8	5.5	2700	6 BD227670	BD227670 Synthesas
44	153.8	5.5	2700	6 AR240705	AR240705 Sequence
45	153.8	5.5	2700	6 AR266980	AR266980 Sequence

ALIGNMENTS

RESULT 1	AF097311	AF097311	2792 bp	mRNA	linear	PLN 22-MAR-2000
LOCUS	Stevia rebaudiana	Stevia rebaudiana	kaurene synthase (KS22-1)	mrna	complete cds.	
DEFINITION	AF097311	AF097311	2792 bp	mRNA	linear	PLN 22-MAR-2000
ACCESSION	AF097311	AF097311	2792 bp	mRNA	linear	PLN 22-MAR-2000
VERSION	AF097311.1	AF097311.1	GI:4959240			
KEYWORDS	Stevia rebaudiana	Stevia rebaudiana				
SOURCE	Stevia rebaudiana	Stevia rebaudiana				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Asterales; Asteraceae; Asteroidae; Eupatorieae; Stevia.					
REFERENCE	1 (bases 1 to 2792)					
AUTHORS	Richman,A.S., Giljzen,M., Starratt,A.N., Yang,Z. and Brandlie,J.E.					
TITLE	Diterpene synthesis in Stevia rebaudiana: recruitment and up-regulation of key enzymes from the gibberellin biosynthetic pathway					
JOURNAL	Plant J. 19 (4), 411-421 (1999)					
MEDLINE	99435880					
PUBMED	10504563					
REFERENCE	2 (bases 1 to 2792)					
AUTHORS	Richman,A.S., Giljzen,M. and Brandlie,J.E.					
TITLE	Direct Submission					
JOURNAL	Submitted (06-OCT-1998) Southern Crop Protection and Food Research Center, Agriculture and Agri-Food Canada, 1391 Sandford Street, London, ON N5V 4T3, Canada					
FEATURES	Location/Qualifiers					
source	1..2792					
	/organism="Stevia rebaudiana"					
	/mol_type="mRNA"					
	/db_xref="taxon:55670"					
gene	1..2792					
	/tissue_type="leaf"					
	/gene="KS22-1"					
	174..2528					
CDS	/gene="KS22-1"					
	/function="synthesis of (-)-kaurene by ionization dependant cyclization of copalyl diphosphate"					
	/note="terpene cyclase"					
	/codon_start=1					
	/product="kaurene synthase"					

/protein_id="A34295.1"
/db_xref="GI:4959241"
/translations="MNLISCIASPLTKSNRPAALSAIHPTASTSHGGQNTPTNLIIDT
TKERIQKFNKIVISVSYDTAWVAMVSPNSPKSPCPPECLANLWLNQNLDSGWL
NHTNHNHPLKDSLSLACIIVALKRWNVGDEQINKGLSFISNLSLASEKQSP9I
GPDIIIPGLLEYAKNLDINLSKQDTPSLMLHRELBOKRCHSNMDGYLAYSIGLG
NLYDMVMVKYQKMGNSVFNPSATAAFINHONPGLNVLNLSLDKFGNAVPTVPH
DLFIRLSWDTIIRLGLSHFRVEIKNVLDETRCWVERDEQIFMDVVTCAAFRLIR
INGVEVSPDPLAETINELALKDEYALLETYHSHLIDYQEDLSGKQLKSADEFKEII
STDSNRLSLIHKREVENALFPINTGLRINTRERNIQLYVNDTRILKTYHSSNTSN
TDYLRILAVEDFYTCQSIYREBELGLERWVENKLDQLKFAKQKTAICYFSAVTLSP
ELSDARISWAKNLDITVDDDFDGGTIDELTNLIQVCEKWNVDYDCSCSEHVRIL
FLAKDAIWIQGEAPKQWQARDYTSHVIOJTWLELMSMLREAIWTRDAYVPTLNEYME
NAVYSFALGIVPEPAIYFVGPKLSEIIVSESEYHNLFKLMSTQGRLLNDLHSPKREPK
EKGKLNVALHLSNGESKVBEEVVEVMMMKKRRKELMKLIIFRENGSIIVRACKDAF
WNWCHVLNFFYANDDGGTNTILDITVDKIIYNPLVLVNEBEOB"

ORIGIN

		Query Match	100.0%;	Score 2792;	DB 8;	Length 2792;		
		Best Local Similarity	100.0%;	Pred. No. 0;				
		Matches 2792;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
Qy	1	CCCACTCATCCTTTATCAACCAATACCATCGTTCGCCACCGGAAGACTGTATCGCAGC	60					
Db	1	CCCACTCATCCTTTATCAACCAATACCATCGTTCGCCACCGGAAGACTGTATCGCAGC	60					
Qy	61	GACCTGACGAGCCCTTTAATCTCTGTTGGACCACTTAACAAATTCGACCATCAGATCT	120					
Db	61	GACCTGACGAGCCCTTTAATCTCTGTTGGACCACTTAACAAATTCGACCATCAGATCT	120					
Qy	121	ACTCCGGTGGACAGTTCATTTGCAATTTAGTAATCAGCAAACTAAACATCAATGAATC	180					
Db	121	ACTCCGGTGGACAGTTCATTTGCAATTTAGTAATCAGCAAACTAAACATCAATGAATC	180					
Qy	181	TTTCACTATGATCGCGTCCCTTTGTTAAACAAATCAAAATCGACCCGGCTCTGTGAG	240					
Db	181	TTTCACTATGATCGCGTCCCTTTGTTAAACAAATCAAAATCGACCCGGCTCTGTGAG	240					
Qy	241	CTATTCATACAGCATCAACTTTCATGTTGCAATTTAGTAATCAGCAAACTAAACATCAATGAATG	300					
Db	241	CTATTCATACAGCATCAACTTTCATGTTGCAATTTAGTAATCAGCAAACTAAACATCAATGAATG	300					
Qy	301	ATACAAACAAAGACCGGATCAAAACACAGTTTAAATAATGTAGAAATTTCTGTTCTTCAT	360					
Db	301	ATACAAACAAAGACCGGATCAAAACACAGTTTAAATAATGTAGAAATTTCTGTTCTTCAT	360					
Qy	361	ATGACACAGCATGGGTAGCCATGGTCCCTTCCTCAAACTACCCAAATCGCCTTTGTTCC	420					
Db	361	ATGACACAGCATGGGTAGCCATGGTCCCTTCCTCAAACTACCCAAATCGCCTTTGTTCC	420					
Qy	421	CTGAGTGTCTCAATTTGGTTAATTAATCAAGTTTAATGATGGTTCATGGGGTCTTGTTTA	480					
Db	421	CTGAGTGTCTCAATTTGGTTAATTAATCAAGTTTAATGATGGTTCATGGGGTCTTGTTTA	480					
Qy	481	ATCACACTCATATAATCAATCAACCGTTTCCTTAAAGATTCCTATCTTCAACATTAGCAT	540					
Db	481	ATCACACTCATATAATCAATCAACCGTTTCCTTAAAGATTCCTATCTTCAACATTAGCAT	540					
Qy	541	GTATTTGGTCATTTAAAGATGGAATGTTGGGGAAGATCAATAAATAAAGGTCTAAGTT	600					
Db	541	GTATTTGGTCATTTAAAGATGGAATGTTGGGGAAGATCAATAAATAAAGGTCTAAGTT	600					
Qy	601	TTATTTGAGTCAAACTCTGTTCCAGTACTGAAAAAAGTCAACCATCTCCCATTTGGTTTG	660					
Db	601	TTATTTGAGTCAAACTCTGTTCCAGTACTGAAAAAAGTCAACCATCTCCCATTTGGTTTG	660					
Qy	661	ACATCATATTTCTGGTTTGGTGTAGTATGCGAAAACTTGGACATAAAACCTCCCTTCAA	720					
Db	661	ACATCATATTTCTGGTTTGGTGTAGTATGCGAAAACTTGGACATAAAACCTCCCTTCAA	720					
Qy	721	AACAAACAGATTTTGTGTTGATGCTACATAAGAGGGAATTCGAGCAAAAAGATGCCATT	780					
Db	721	AACAAACAGATTTTGTGTTGATGCTACATAAGAGGGAATTCGAGCAAAAAGATGCCATT	780					

Qy	781	CBAATGAGATGGATGGATACCTTGGCGTATATCTCTGAAGGACTCGGTAAATTTATATGATT	840
Db	781	CAAATGAGATGGATGGATACCTTGGCGTATATCTCTGAAGGACTCGGTAAATTTATATGATT	840
Qy	841	GGAAATGGTGAAGAAATATCAGATGAAAAATGGTCTCTGTTTCAACTCACCATCAGCAA	900
Db	841	GGAAATGGTGAAGAAATATCAGATGAAAAATGGTCTCTGTTTCAACTCACCATCAGCAA	900
Qy	901	CAGCTGCTGCTTTCATTAATCATCAAAATCCTGGTGTCTTAATTTTAAATTCACATTT	960
Db	901	CAGCTGCTGCTTTCATTAATCATCAAAATCCTGGTGTCTTAATTTTAAATTCACATTT	960
Qy	961	TGGACAAAGTTTGGTAAATGCGAGTCCCAACAGTTTATCTCATGATTTATTTATCCGACTTT	1020
Db	961	TGGACAAAGTTTGGTAAATGCGAGTCCCAACAGTTTATCTCATGATTTATTTATCCGACTTT	1020
Qy	1021	CTATGGTTGACAAATGAAAGATTAGGAATTTTACACCAATTCAGAGTGAATTAATAA	1080
Db	1021	CTATGGTTGACAAATGAAAGATTAGGAATTTTACACCAATTCAGAGTGAATTAATAA	1080
Qy	1081	ATGTTTATGATGAAACATACAGATGTTGGGTGGACGAGATGACCAATATTCATGATG	1140
Db	1081	ATGTTTATGATGAAACATACAGATGTTGGGTGGACGAGATGACCAATATTCATGATG	1140
Qy	1141	TTGTAAACATGCTGCTTTAGCCCTTTCCGTTATTAAAGGATCAATGGGTATGAAGTTTCCCAG	1200
Db	1141	TTGTAAACATGCTGCTTTAGCCCTTTCCGTTATTAAAGGATCAATGGGTATGAAGTTTCCCAG	1200
Qy	1201	ATCCATTTGGCTGAAATTAATTAAGATTTAGCTTTGAAAGACGAATATGCGAGCTCTTGAAA	1260
Db	1201	ATCCATTTGGCTGAAATTAATTAAGATTTAGCTTTGAAAGACGAATATGCGAGCTCTTGAAA	1260
Qy	1261	CATATCATGCGTCAATATATTAACCAAGAGGATTTATCTCTCGGAAAAACAAATCTTGA	1320
Db	1261	CATATCATGCGTCAATATATTAACCAAGAGGATTTATCTCTCGGAAAAACAAATCTTGA	1320
Qy	1321	AGTCAGCTGATTTCTCTCAAGAGATAATATCCACTGATTTCAAAACAGGCTTTCTAAATTA	1380
Db	1321	AGTCAGCTGATTTCTCTCAAGAGATAATATCCACTGATTTCAAAACAGGCTTTCTAAATTA	1380
Qy	1381	TTCAAAAGAGGTGGAAAAATGCTTAAAGTTCCCTATCAATACCGGTTTGAAGCGCATAA	1440
Db	1381	TTCAAAAGAGGTGGAAAAATGCTTAAAGTTCCCTATCAATACCGGTTTGAAGCGCATAA	1440
Qy	1441	ACACTAGACGAATATACAGCTTTTACAAATGAGCAATATACAAAGATTTCTGAAAACTACAT	1500
Db	1441	ACACTAGACGAATATACAGCTTTTACAAATGAGCAATATACAAAGATTTCTGAAAACTACAT	1500
Qy	1501	ATCACTCATCAAAATATTAGTAACACTGATTAACCTAAGGTTGGCTGTTGAAGATTTCTACA	1560
Db	1501	ATCACTCATCAAAATATTAGTAACACTGATTAACCTAAGGTTGGCTGTTGAAGATTTCTACA	1560
Qy	1561	CCTGCCAATCTATTATTCGTGAAGAAATTAAGGCTTTGAAGGTGGGTGGGTAGAGAATA	1620
Db	1561	CCTGCCAATCTATTATTCGTGAAGAAATTAAGGCTTTGAAGGTGGGTGGGTAGAGAATA	1620
Qy	1621	AGTTGGACCACTCAAGTTTCTAGGCAAAAGACCGCTACTGTTATTTCTCTGTTGCTG	1680
Db	1621	AGTTGGACCACTCAAGTTTCTAGGCAAAAGACCGCTACTGTTATTTCTCTGTTGCTG	1680
Qy	1681	CAACACATTTCTCTCCGAAATTAATCAGATGCGCTATTTTCATGGGCCAAAAATGGCATAT	1740
Db	1681	CAACACATTTCTCTCCGAAATTAATCAGATGCGCTATTTTCATGGGCCAAAAATGGCATAT	1740
Qy	1741	TAACTACAGTAGTGTGATGACCTTTTGTATATCGGTGTACAAATCGATGAATTTGACCAACC	1800
Db	1741	TAACTACAGTAGTGTGATGACCTTTTGTATATCGGTGTACAAATCGATGAATTTGACCAACC	1800
Qy	1801	TGATTTCAATGTTGAAAAATGGAATGTAGATGTGCAACAGGATTTGTTGTTTTCAGAGCATG	1860
Db	1801	TGATTTCAATGTTGAAAAATGGAATGTAGATGTGCAACAGGATTTGTTGTTTTCAGAGCATG	1860

Db 706 CATGGTCCCTTCCCAACTCACCCAAATCGCCTTGTTCCTCGAGTGTCTCAATTGGTT 765
Qy 440 AATTAATAATCAGCTTAATGANGTTCATGGGGTCTTGTGTAATACACACTCATTAATCATAA 499
Db 766 AATTAATAATCAGCTTAATGANGTTCATGGGGTCTTGTGTAATACACACTCATTAATCATAA 825
Qy 500 TCACCCGTTGCTTAAAGATTCCTCATCTCAACATTAAGCATGTATGTTGCAATTAATAAG 559
Db 826 TCACCCGTTGCTTAAAGATTCCTCATCTCAACATTAAGCATGTATGTTGCAATTAATAAG 885
Qy 560 ATGGAATGTTGGGGAAGATCAAAATAAATAAAGTCTAAGTTTATTTAGTCAAACTTTGC 619
Db 886 ATGGAATGTTGGGGAAGATCAAAATAAATAAAGTCTAAGTTTATTTAGTCAAACTTTGC 945
Qy 620 TTGAGTACTGAAAAAGTCAACCATCTCCATTTGGTTTGGACATCATATTTCCCTGGTTT 679
Db 946 TTGAGTACTGAAAAAGTCAACCATCTCCATTTGGTTTGGACATCATATTTCCCTGGTTT 1005
Qy 680 GCTTGAGTATCGAAAAAATTTGGACATAAACCCTCTTTCAAAACAAAACAGATTTAGTTT 739
Db 1006 GCTTGAGTATCGAAAAAATTTGGACATAAACCCTCTTTCAAAACAAAACAGATTTAGTTT 1065
Qy 740 GATGCTACATAAGAGGAAATTTGGAGCAAAAAGATGCCATTTCAAATGAGATGCGATGATA 799
Db 1066 GATGCTACATAAGAGGAAATTTGGAGCAAAAAGATGCCATTTCAAATGAGATTTGATGATA 1125
Qy 800 CTTGGCGTATATCTCTGAAGACTCGGTAAATTTATATGATGGAAATGTTGGAGAAATA 859
Db 1126 CTTGGCGTATATCTCTGAAGACTCGGTAAATTTATATGATGGAAATGTTGGAGAAATA 1185
Qy 860 TCAGATCAAAAATGCTTCTGTTTCACTCAACATCAGCAACAGCTGCTGCTTTCAATTA 919
Db 1186 TCAGATCAAAAATGCTTCTGTTTCACTCAACATCAGCAACAGCTGCTGCTTTCAATTA 1245
Qy 920 TCATCAAAAATCCTGGTGTCTTAATATTTAAATTCATTTTGGACAAAGTTTGGTAATGC 979
Db 1246 TCATCAAAAATCCTGGTGTCTTAATATTTAAATTCATTTTGGACAAAGTTTGGTAATGC 1305
Qy 980 AGTCCCAACAGTTTATCCTCATGATTTATTTATCGGACTTCTATGTTGACACAATTA 1039
Db 1306 AGTCCCAACAGTTTATCCTCATGATTTATTTATCGGACTTCTATGTTGACACAATTA 1365
Qy 1040 AAGATTAGCAATTTACACACATTTTCAAGTGGAAATTTAAATGCTTTTGTAGATGAAACATA 1099
Db 1366 AAGATTAGCAATTTACACACATTTTCAAGTGGAAATTTAAATGCTTTTGTAGATGAAACATA 1425
Qy 1100 CAGATGTTGGTGGAAACAGATGAGCAAAATTTATGATGATTTGTAAATGCTTTTATG 1159
Db 1426 CAGATGTTGGTGGAAACAGATGAGCAAAATTTATGATGATTTGTAAATGCTTTTATG 1485
Qy 1160 CTTTCCGTTTATTAAGATCAATGGGTATGAAGTTTCCACAGATCCATTTGGCTGAAATTA 1219
Db 1486 CTTTCCGTTTATTAAGATCAATGGGTATGAAGTTTCCACAGATCCATTTGGCTGAAATTA 1545
Qy 1220 TAATGAATTTAGCTTTGAAAGCAATATGCAAGTCTTTGAAACATATCATGCTGAT 1279
Db 1546 TAATGAATTTAGCTTTGAAAGCAATATGCAAGTCTTTGAAACATATCATGCTGAT 1605
Qy 1280 ATTATACAAAGGATTTATCTTCTGGAAACAAATCTTGAAGTCAAGTCAAGTCAAGT 1339
Db 1606 ATTATACAAAGGATTTATCTTCTGGAAACAAATCTTGAAGTCAAGTCAAGTCAAGT 1665
Qy 1340 AGAGATATATCCATGATTCACACAGCTTTTCTAAATTAATTTCAAAAGAGGTGGAAAA 1399
Db 1666 AGGATATATTCCTACTGATTCACACAGCTTTTCTAAATTAATTTCAAAAGAGGTGGAAAA 1725
Qy 1400 TGCTCTTAAGTTCCCTATCAATACCGGTTTGAAGCGCATATAAACACTAGACGAAATATACA 1459
Db 1726 TGCTCTTAAGTTCCCTATCAATACCGGTTTGAAGCGCATATAAACACTAGACGAAATATACA 1785
Qy 1460 GCTTTACATGTAGACAAATCAAGATTTCTGAAACTACATATCACTCATCAATATTAG 1519

Db 1786 GCTTTACAATGTAGACAAATCAAGAAATTTGAAAACTACATATCACTCATCAAAATATTAG 1845
Qy 1520 TAACACTGATTAACCTAAGGTTGGCTGTTGAAGATTTCTACACCTGCGCAATCTATTATTCG 1579
Db 1846 TAACACTGATTAACCTAAGGTTGGCTGTTGAAGATTTCTACACCTGCGCAATCTATTATTCG 1905
Qy 1580 TGAAGAATTAATAAGGTTCTTGAAAGGTTGGTGTAGAGAAATAAGTTGACACAGCTCAAGTT 1639
Db 1906 TGAAGAATTAATAAGGTTCTTGAAAGGTTGGTGTAGAGAAATAAGTTGACACAGCTCAAGTT 1965
Qy 1640 TGCTAGCAAAAAGACCGCTTACTGTTATTTCTGTTGCTGCAACACTTTGCTGCTCCGA 1699
Db 1966 TGCTAGCAAAAAGACCGCTTACTGTTATTTCTGTTGCTGCAACACTTTGCTGCTCCGA 2025
Qy 1700 ATTATCAGATGCGGTATTTTCATGGGCCAAAAATGGCATATTAACACTAGTAGTTGATGA 1759
Db 2026 ATTATCAGATGCGGTATTTTCATGGGCCAAAAATGGCATATTAACACTAGTAGTTGATGA 2085
Qy 1760 CTTTTTTGATATCGGTGGTACAATTCGATGAATTGACCAACTGATTCATGATTCATGTTGAAAA 1819
Db 2086 CTTTTTTGATATCGGTGGTACAATTCGATGAATTGACCAACTGATTCATGATTCATGTTGAAAA 2145
Qy 1820 ATGGAATGTAGATGTGCAAGGATTTGTTGTTTCAAGCATGTTCCGATTTTATTTTATG 1879
Db 2146 ATGGAATGTAGATGTGCAAGGATTTGTTGTTTCAAGCATGTTCCGATTTTATTTTATG 2205
Qy 1880 ATTAAGAGATGCAATCTGTTGGATTTGAGATGAAGCTTTTAAATGGCAAGCGCGCGATGT 1939
Db 2206 ATTAAGAGATGCAATCTGTTGGATTTGAGATGAAGCTTTTAAATGGCAAGCGCGCGATGT 2265
Qy 1940 AACTAGCCATGTTTATCAAACTTGGTTGGAACTTAATGAATGATGTTGAGAGAGCTAT 1999
Db 2266 AACTAGCCATGTTTATCAAACTTGGTTGGAACTTAATGAATGATGTTGAGAGAGCTAT 2325
Qy 2000 ATGCAAGAGATGCTTATGTCGCAACATTAATGAATATATGAAACCGTTTACGCTGTC 2059
Db 2326 ATGCAAGAGATGCTTATGTCGCAACATTAATGAATATATGAAACCGTTTACGCTGTC 2385
Qy 2060 ATTTGCAATTAGCCCGGATTTGCAAGCGGCTATTTACTTTTGTGGGGCCCAAAATATCAGA 2119
Db 2386 ATTTGCAATTAGCCCGGATTTGCAAGCGGCTATTTACTTTTGTGGGGCCCAAAATATCAGA 2445
Qy 2120 GGAGATTTGTAAGGCTCTGAATATCATATCTAATTTAAGTAATGAGCAGCGAGGTCG 2179
Db 2446 GGAGATTTGTAAGGCTCTGAATATCATATCTAATTTAAGTAATGAGCAGCGAGGTCG 2505
Qy 2180 ACTTTTAAACGATATCCATAGCTTCAAGAGGGAATTTTAAAGCAAGCAAAATTAACCGGCT 2239
Db 2506 ACTTTTAAACGATATCCATAGCTTCAAGAGGGAATTTTAAAGCAAGCAAAATTAACCGGCT 2565
Qy 2240 AGCAATTCATTTGAGTAACCGAGAAAGTGGGAAAGTGGAAAGAGGTTGTGGAGGAGAT 2299
Db 2566 AGCAATTCATTTGAGTAACCGAGAAAGTGGGAAAGTGGAAAGAGGTTGTGGAGGAGAT 2625
Qy 2300 GATGATGATGATTAATAAACAAGAGGAAGATTAATGAATTAATTTTGAAGAAATG 2359
Db 2626 GATGATGATGATTAATAAACAAGAGGAAGATTAATGAATTAATTTTGAAGAAATG 2685
Qy 2360 TAGCATTTCTTCTAGAGCTTGAAGATGCAATTTTGGAAACATGTCACGTTTGAATTT 2419
Db 2686 TAGCATTTCTTCTAGAGCTTGAAGATGCAATTTTGGAAACATGTCACGTTTGAATTT 2745
Qy 2420 TTTTATACGCAACGATGACGGTTTACTGGGAAACACGATTTCTTGATATCTGTGAAGGACAT 2479
Db 2746 TTTTATACGCAACGATGACGGTTTACTGGGAAACACGATTTCTTGATATCTGTGAAGGACAT 2805
Qy 2480 CATTTACAAACCGTTGCTTGTGATGAAATGAAGCAAAAGTAATTTGGCTTTAA 2539
Db 2806 CATTTACAAACCGTTGCTTGTGATGAAATGAAGCAAAAGTAATTTGGCTTTAA 2865
Qy 2540 AAGATGATTAATAATATCATGCTTGTGACGGGGTACTGTTGTAGTTGTATATAAAGGTT 2599
Db 2866 AAGATGATTAATAATATCATGCTTGTGACGGGGTACTGTTGTAGTTGTATATAAAGGTT 2925

```
QY 2600 GTAGTTGTTATATAAAGGTAATAGGTAATCAATAGAAAGCTTAAGTTATTAAGTTTTTCC 2659
Db 2926 GTAGTTGTTATATAAAGGTAATAGGTAATCAATAGAAAGCTTAAGTTATTAAGTTTTTCC 2985
QY 2660 CTCGTGTCAACACCGGTGAGGTTCTTTGTTAAAGCAGCTTTATTTATGTTTGTAGTACGACAC 2719
Db 2986 CTCGTGTCAACACCGGTGAGGTTCTTTGTTAAAGCAGCTTTATTTATGTTTGTAGTACGACAC 3045
QY 2720 TTGGTGGTTGTTGGTTGTTTATTTCTTAAGAGGCTTTCATTGTTCTTTAAAAAAA 2776
Db 3046 TTGGTGGTTGTTGGTTGTTTATTTCTTAAGAGGCTTTCATTGTTCTTTAAAGAA 3102

RESULT 3
LOCUS AB031205
DEFINITION Lactuca sativa Lsks1 mRNA for ent-kaurene synthase N01, complete cds.
ACCESSION AB031205
VERSION AB031205.1 GI:9971224
KEYWORDS ent-kaurene synthase N01.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE 1 (bases 1 to 2594)
            Toyomasu, T., Mitsuhashi, W. and Kamiya, Y.
            Gibberellin biosynthetic enzyme
            Published Only in DataBase (2000)
REFERENCE 2 (bases 1 to 2594)
            Toyomasu, T., Mitsuhashi, W. and Kamiya, Y.
            Direct Submission
            Submitted (17-AUG-1999) Tomonobu Toyomasu, Yamagata University,
            Faculty of Agriculture, Wakaba-cho 1-23, Tsuruoka, Yamagata
            997-8555, Japan (E-mail: toyomasu@tds1.tr.yamagata-u.ac.jp,
            Tel:81-235-28-2861, Fax:81-235-28-2812)
FEATURES             Location/Qualifiers
     source            1..2594
                        /organism="Lactuca sativa"
                        /mol_type="mRNA"
                        /strain="Grand Rapids"
                        /db_xref="taxon:4236"
     gene              1..2594
                        /gene="LsKS1"
     CDS               87..2453
                        /product="ent-kaurene synthase N01"
                        /protein_id="BAB12441.1"
                        /db_xref="GI:9971225"
                        /translation="MNIAQITSSAMLVPSHPIHRSWVNCMVQNPGLRTASSQA
                        GQNPVTMTLVTKERIKLPNNVEVSSYDTAWVAMPSPKSPCFDCLNWL
                        DNQLDGSGLVPHOSPLIKDLSLTACVLAKRWNVGDKQINKGLHYIESNFASVT
                        DKNQASPFQDIIIFQMLBYAKDLIKLPLNOLHLSVLMHERELELRCHSNGREAYL
                        AVISRLGNLNDNMVQKMGSLFNSPSATSVLIHQHAGCLHYLTSLDKFEN
                        AVPTYPIDLYRLSNVDLERLGIKHPWVEIQNVLDIETRCWQGVQVDFMDVVTC
                        ALAFVRLSNGVESSDPLAKITKBDYNNSPKPFKDVYTSLEVYKASQIIYQBELA
                        FREQNLTSYPSNLSNLYLKEVDLALFPFNGSLRMSRTRNEHYNLNHLRIKLT
                        TYSSNINSMKILKLAQVDFNQCQIYCEELKDLRWVVENLRLDKLPAROKTAYCF
                        SAASFLSPDLKADVQNSWAKSILITVDFDFVGSMDLNVFNVHIIIEKNVNVEND
                        CCSSEVGLVFLAKDAVCGIDKAFKIQERNITSHVIELWLDVKSMLREALWAKDGS
                        IPTFNEYMNGVSVFALGVIPVLPFIYFLGKLSSEVVQSSEYHKLYEVNSTQRLMND
                        IHSFKRKAGKGLNALVNSDKGSVEEVVEEMKILITKSORKEMKMLVLETRGSV
                        VPRVCKDVFNNVNCVNLNLFYATDGDFTGNAILDVVKELIIEYEPVSEHLL"

Query Match          51.2%;   Score 1428.6;   DB 8;   Length 2594;
Best local Similarity 75.8%;   Pred. No. 9.5e-270;
Matches 1833;   Conservative 0;   Mismatches 529;   Indels 57;   Gaps 3;

ORIGIN
```

```
1255 AGTTAGCTTTTAGAGACAAAACCTTGA-----CAAGTTTATCTCC 1293
1354 CTGATTCGAACAGGCTTTCTAAATTAATTCACAAAGAGGTGGAAAATGCTCTTAAAGTCC 1413
1294 CTTCAAGTAAACAGCTCTCTAAATATATATCTCTCAAGAGAGGTGGATGATGCTCTTAAAGTCC 1353
1414 CTATCAATACCGGTTTAGAAGCGAATAAACAACACTAGAACGAATATATACAGCTTTTACAATGTAG 1473
1354 CTTTAAATGGAGAGCTTAGAAGCGAATGTCCTACTAGAGAAACATAGAGCAATTCACAAATTAA 1413
1474 ACAATACAGAATTTCTGAAAACCTACATATCACTCATCAAAATATTAGTAACAACCTGATTACC 1533
1414 ACCATACAGAATCTTTGAAAACCTACATATAGCTCATCAAACTATTAGCAACAAGATTACC 1473
1534 TAAGTTGGCTGTTGAAGATTTCTACACCTGCCAATCTATTATTCGTAAGAAATTAAG 1593
1474 TAAAGTTAGCTGTGCAAGATTTCAATGAATGCCAATCTATCTATTGTGAAGAACTAAAG 1533
1594 GTCTTTGAAAGTGGTGGTAGAGAAATAAGTTGGACCACTCAAGTTTGTAGGCAAAAGA 1653
1534 ATCTTGAAGGTGGTGGTAGAGAAACAGATTAGCAAGCTCAAAATTTGCTTAGACAGAGA 1593
1654 CCGCTACTGTTATTTCTGTTGCTGCAACACTTTCTGCTCTCCGAAATTTATCAGATCGC 1713
1594 CAGCTACTGTTACTTTCTGCTGCTTCAATTTCTTCTTCTCCCTGATTTATCAGATGCC 1653
1714 GTATTTTCATGGCCAAAATGGCAATTAATACTAGTAGTAGTAGTGTATTTTTCATATCG 1773
1654 GAATTTTCATGGCCAAAATGAGTACCTCTCACTAGCAATTAATGATGACTTTTTTGTATGTTG 1713
1774 GTGTACAATCGATGAATTTGACCAACCTGATTTCAATGTTTGAATAATGGAATCTAGATG 1833
1714 GAGATCTATGGAAGACTGGTGAACCTTTGTTTCACTATTTGAAAATGGAATGTAACG 1773
1834 TCGAACAGGATTTGTTTCAGAGCAATGTTGCGATTTTATTTTATAGCATTTAAAGATGCAA 1893
1774 TTGAAAATGATTTGCTGTTTCAGAGAAATGTCGCTTTTATTTTATAGCACTTTAAGGATGCG 1833
1894 TCTGTTGGAATGGAGATGAGCTTTTAAATGCGACGCGCGATGTAACCTAGCATGTTA 1953
1834 TCTGTTGGAATGGAGACAAAGCTTTTAAATGATTCAGAAACGCAATATACAGACCATGTTA 1893
1954 TTCAAACTTGTGTCGAACTAATGATATGATGTTTGCAGAGAGCTATATGACACAGAGATG 2013
1894 TTGAAATTTGTTGGAATTTGTTGGAAGAGTATGTTGAGAGAGCTATATGCGCGAAGATG 1953
2014 CTTATGTGCCAACATTAATGAATATATGGAACCGCTTAGTGTCTATTTGCATTAGGCC 2073
1954 GTTCAATTTCCAAACAATAAATGAATATATGGAATAATGTTAGTGTCTATTTGCTTTAGGCC 2013
2074 CGATTGTCAAGCGCGCTATTTACTTTTGTGGGCGCCAAATATCAGAGAGATTTGTGAAA 2133
2014 CGATTGCTCCCTACTCTTTACTTTTATGAGGTTTAAATTTGTCGAGGAAGTTGTTGAGA 2073
2134 GCTCTGAATATCATATCTATTTAAGCTAATAGCAACGCGAGGTGCGACTTCTAAACGATA 2193
2074 GCTCCGAGTATCAAGACTTTATGAGGTTATGAGCACTTCAGCGTCTCTCATGAATGATA 2133
2194 TCCATAGCTTCAAGAGGAATTTAAGAGAGCAATTAACCGCGGTAGCATTTGATTGA 2253
2134 TCCATAGCTTCAAGAGGAAAAGAGCGCGGGAATTTGAATGCTGTGGCAATTTGATACGA 2193
2254 GTAACCGGAAAAGTGGGAAAAGTGAAGAGAGGTTTGTGGAGAGATGATGATGATTA 2313
2194 GTGATGGAAGAGTGGGAGTGTGGAAGAGAGGTTTGTAGAGAGATGAAGATTTTGACA 2253
2314 AAAACAAGAGAAAAGAAATTAATGAATTTAATTTTGAAGAAAATGTTAGCATTTGTTCTTA 2373
2254 AAAGTCAAGAGGAAAGAGATGATGAAATTAGTTTGTGAAAACAAAAGGAAGTGTGTTCCAA 2313
2374 GAGCTTTGAAGATGCAATTTTGGACAGTGTGTCAGGTGTTGAAATTTTATTTTACCAACG 2433
```

```
2314 GAGTATGCAAGAGATGATTTTGGAAACATGTGCAATGTGTGAATCTTTTTTATGCGACAG 2373
2434 ATGACGGGTTTACTTGGAAACACGATTTCTTGATACCTGTGAAGGACATCATTTTACAACCGT 2493
2374 ATGATGGGTTTCACTTGGAAATGCGCATTTCTTGATGTTTGAAGGAATTTATATGAACCTG 2433
2494 TGTGCTTGTGCAATGAAATGAAGAACAAAGGTAATTTGGCCTTTTAAAGATGATAAATAAT 2553
2434 TGTCCCATGAGTTGATATAATCATTTGTTATACCAATTCGTATTATATTATTAGCATCTC 2493
2554 ATCATGCTCTTGAAGGGGGTACTGTTGTAGTTGTATATAAAGGTTGATGTTATATAATA 2613
2494 AAAAATGAACACAGCATACTAGTATTCTAGATCTACTCATGTTTATTATTATTATTTTTT 2553
2614 AAGGTAATAGTGAATCAAT 2632
2554 AATATGTTTGTAGCTAT 2572

RESULT 4
AB045310
LOCUS AB045310 2450 bp mRNA linear PLN 06-JUN-2002
DEFINITION Cucumis sativus CsKS1 mRNA for ent-kaurene synthase, complete cds.
ACCESSION AB045310
VERSION AB045310.2 GI:21326756
KEYWORDS Cucumis sativus (cucumber)
SOURCE Cucumis sativus
ORGANISM Cucumis sativus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
REFERENCE 1
AUTHORS Shirai,K., Fujino,K. and Masuda,K.
TITLE Cloning and sequencing of an ent-kaurene synthase cDNA from cucumber
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2450)
AUTHORS Shirai,K., Fujino,K. and Masuda,K.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-2000) Kikuko Shirai, Graduate School of
Agriculture, Hokkaido University, Laboratory of Horticulture
Science, Kitaku Kitag nishi9, Sapporo, Hokkaido 060-8589, Japan
(E-mail:kikou@res.agr.hokudai.ac.jp, Tel:81-11-706-3875(ex.3875),
Fax:81-11-706-4937)
COMMENT On Jun 5, 2002 this sequence version replaced gi:11862932.
FEATURES
Location/Qualifiers
1..2450
/organism="Cucumis sativus"
/mol_type="mRNA"
/db_xref="taxon:3659"
/clone="CsKS1"
1..2450
/gene="CsKS1"
23..2380
/gene="CsKS1"
/codon_start=1
/product="ent-kaurene synthase"
/protein_id="BAB19275.1"
/db_xref="GI:11862933"
/translation="MNLSPRTNLGCTASSASLFGLDVGYTKTKTGALPEETKERI
KKLFWNELSISAYDTAWAVNVPNSLNKPLPPECINWLDHNPQSGWGLHDHQL
VMKATLSTLACVLTLEKRWIDIGDDHMSKALSFIKSNATASDENQSPFVDIIFPGM
IETAKDLNLNPLASNNVDALVQKKELELRSCSNSEGGKAYLAYAIVPDIYARLH
VMRYORNGSLFSPSTTAVAFMRNDGCFNYLRSVLQKPFHSPVPAIYPLDIYARLH
MVDLQKLGIDGHFKDRIRSLDVTYSCMWGENTIFLDASTCMAFMRLRVEGYDVS
SDQLTORSEGLFNCCLGGHLKDFSALEFKASOLIIYPDEFILLENINSTRFLNHG
LNSGSHRTERLRLVKOEAVNAFEPPNLSRLERLSNRKALESYSGDILVRSKTAACI
NFGHQDFLELAVEDNTLQGHRRKELKELKWLTKDKLKFARQKLAVRCYFSAAT
LITPELCDARLSWAKNGVLTVDVDFDVGSEELVNLILQVEKWDASGTGYCSKE
VEIIFLALHSTICEIGKKALFWQGRSVNRVIDIWLALLEMRKEALKNKVVPSLD
EYEMNGVVALGPVLPVTFYVFGPKLPEEIVNCEYKCLKFKLMSTSGRLNDRTFD
RESSEGNLALSLYMI SAGGLTKEATEAMKGDVDRTRRELLRLVLQENSTIPRACK
DLFWKSCVHLFRKDDGTSHELMSAKALFEQPMVLDELNKK"
```

ORIGIN

Query Match 29.0%; Score 809.4; DB 8; Length 2450;
Best Local Similarity 62.0%; Pred. No. 2.1e-148;
Matches 1410; Conservative 0; Mismatches 811; Indels 54; Gaps 6;

QY	267	GGTGGCAAACTAATCCCACTAATCTGATCATGTACAACCAAGAAGCGATCCAAAA	326
DB	101	GGCACAAAACATAAACTGGAGCTTTGCCTTTGAAGAAACAAAAGAAGATTA AAAA	160
QY	327	CAGTTTTAAAAATGTAGAAATTTCTGTTTCTTCATATGACACAGCATGGGTAGCATTG	386
DB	161	TTGTTCAAATAAGTTGAACCTTCAAATTTGTCATATGACTGCAATGGGTGCAATGTC	220
QY	387	CCTTCTCCAAACTCACCAAAATCGCCTGTTCCCTGAGTGCTCAATTGGTTAATTAAT	446
DB	221	CCTTCTCCAAACTCTCTTTAATAAACCTCTTTTTCCTGAGTGCAATAAAGCTGGGTAT	280
QY	447	AATCAGCTTAATGATGGTTTCATGGGTCCTGTTAAACACATCAATAATCATAATCACCG	506
DB	281	CATCAAAACCCGTGATGGGTCAATGGG-----CATACTCATGACCATCGTTG	328
QY	507	TTGCTTTAAAGATTCCTTATCTTCAACATTAGCATGTATTTGTGCATTA AAAAGAT	566
DB	329	GTCATGAAGGCCACTCTCTTATCCAATAGCATGTCTTACTCTTAAGCATGGGAT	388
QY	567	GTTCGGGAAGATCAAAATAAATAAGGTCCTAAGTTTTATTGAGTCAAAATCTCGCTCAGCT	626
DB	389	ATCGGTGATGATCATATAGCAAGGCCCTTAGTTTTATCAAGTCTAATATAGCTTCAGCT	448
QY	627	ACTGAAAAAGGTAAACCATCTCCCATTTGGTTTGGACATCATATTTCCCTGGTTGCTTGAG	686
DB	449	ACTGATGAGAATCAACGCTTCCTCTGTGGGATTTGACATAATTTTCCCTGGTATGATGAG	508
QY	687	FATCGGAAAACTTTGGACATAAAACCTCCTTTCAAAA CAAAACAGATTTTAGTTTGATGCTA	746
DB	509	FATGCTAAAGACTTGAATTTGAAATCTACCTTGGCATCAATGAATGTGGATGCTTTGGTT	568
QY	747	CATAAGAGGGAATTTGGAGCAAAAAGAT-----GCCATTTCAAATGAGATGGATGANTAC	800
DB	569	CAAAAGAAAGATTTGGAGCTTTAGAGCTGCTGTAGCAACTCTGAAGGAGAAAAGCCTAT	628
QY	801	TTGGCGTATATCTCTGAAGGACTCGGTAAATTTATATGATTTGGATATCGTGAAGAATAT	860
DB	629	TTAGCGTATGTTTCAGAAGGAATTTGGAATAATTACAGGACTGGGAATTTGGTCATGCGATAT	688
QY	861	CAGATCAAAAATGGTTCGTTTTCBAACTCACCATCAGCAACAGCTGCTGCTTTCAITTAAT	920
DB	689	CAAAAGGAAGCGGATCACTGTTTAGTTCTCANCCACACGGCAGTGGCTTTATGCGAC	748
QY	921	CATCAAAATCCCTGGTGTGCTTAAATTAATTAATTAATTCATTTTGGCAAGTTTGGTAATGCA	980
DB	749	AGAAATGATATGGGCTGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	808
QY	981	GTCCCCAACAGTTTATTCCTCATGATTTATTTATTCGACTTTCTATGTTTGGACACAAATTGAA	1040
DB	809	GTTCCCGCAATATATCCTCTTGATATATANGCTCGTTTTCACATGCTGGTTGATAGCCTTCAA	868
QY	1041	AGATTAGGAATTTCAACCAATTTCAAGTGGGAATTTAAAAATGTTTTAGATGAAACATAC	1100
DB	869	AAACTGGGATTTGANGCCATTTCAAAGATGAGATTAGAGTGTATTATGATGAACAATAC	928
QY	1101	AGATGTTGGTGAACGAGATGAGCAAAATATTTCATGGATGTTTGAACATGTGCTTTAGCC	1160
DB	929	AGCTGTTGGAATGCAAGGGAGGAACAATATTTCTAGATGCTTCAACTTTGTGCAATGGCC	988
QY	1161	TTTCGGTATTAATAAGGATCAATGGGTATGAAGTTTCCCAGATTCOAATGGCTGAAATTAAT	1220
DB	989	TTCCGGATGTTACGTGTGAAGGATATGATGTTTCTTCAGATCAATTCGACTCAATTTCTCA	1048
QY	1221	AA-----TGAAATAGCTTTCAAGACGATATAGCACTCTTGAA	1259
DB	1049	GAAGGTCCTCTTTTCAAATTTGCCTCGAGAGCAATTAAGAATTTTAGTGCTCACTAGAG	1108


```

rosids; eusoids I; Cucurbitales; Cucurbitaceae; Cucurbita.
1 (bases 1 to 2658)
Yamaguchi,S., Saito,T., Abe,H., Yamane,H., Murofushi,N. and
Kamiya,Y.
Molecular cloning and characterization of a cDNA encoding the
gibberellin biosynthetic enzyme ent-kaurene synthase B from pumpkin
(Cucurbita maxima L.)
Plant J. 10 (2), 203-213 (1996)
9637664
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1..2658
/organism="Cucurbita maxima"
/mol_type="mRNA"
/db_xref="taxon:3661"
/tissue_type="immature seeds"
138..2507
/function="ent-kaurene synthase B in gibberellin
biosynthesis"
/note="terpene cyclase"
/codon_start=1
/product="ent-kaurene synthase B"
/protein_id="AAB39482.1"
/db_xref="GI:1431870"
/translation="NVLSPPTGVARFAASSSSSSASLPFGVDVDTTTKTGALHFEEET
KERIKKGLBELSVSYADTAWVAVPSPNSLNPLFPECINWLDSHADGSGWLLH
NDQDLKANLSTLASVCNLQPLKRWNIAGHDHMSXALDFIKSNIASATDENORSPVGDDI
IPGMTEYAKDNLNLEPLAPTNIDLVRLKEBELRSCRSNSEGGKAVLATVSEGIKGLQ
DMDVMYOKRGSLFNFSSTTTAAAFMRHENDDCEDLYLSLLQKFQDGSVPTIYPDDIY
ARLHWDSLQKFGIARHPKEEISRVILDETYRCMWQGEENIFLDASTCAWAFLMLRVEG
AYLHVSQDLTFQSEDIFPNCLIGLVKDFGASLEYKASQIIITHPDESVLKNINSWTGRF
LKHGUSDSDSRDTSVVQZAVNALRPYNATLERLIISKRAEMESYGDDIVRIKSPP
YAACINFPHGDPLELAVEDNTLQRHLKLEBELQRWVENKLDELFEFRHLGCYCVA
AACTINDPELHDARTAWAQNGVLTITVVDDFDGGSSEELDNLIIEVKWDPPGEGVY
CSKDVEIYEAHLHSVTCEIRALVQWGRSMVRNVIDGMALLKWMRKREABWSNKNDI
PSMGWEAOHYSFALGPDIPLMPFVGPKLSEMI GSCYEYQKLYKLMSTAGLRKNDI
RSYDRCKEGKLNILLMWIDGGGNVTKEEATEAIKGDFERAIRELLGLVLENTTPIP
RACKDLFWKLMISVINLFYMEDDGYTSNRLMNTVKAMFEQPMDDLDAALK"
```

1616	GAATTAAGTTGGACCGACTCAAGTTTGTCTAGCAGAAAGACCGCCTACTCTGTTATTTCTCTGT	1675
Qy		
1604	AAACAAATTTGGACGAGTTGAAATTTTTCAGACTGCACCTAGGGTACTGCTATTTTGTCTGC	1663
Db		
1676	TGCTGCAACACATTTCTCGTCTCCCGAATTATACAGATGCGCGTATTTTCATGTGGGCCCAAAATGG	1735
Qy		
1664	GGCAGCGCCCTTACTGATCTCTGAACTTCATGATGCTCGCATAGCATGGGCACAAAATGG	1723
Db		
1736	CATATTAACTACAGTAGTTGATGACTTTTTCATATCGGTGCTCAATTCGATGAATTCGAC	1795
Qy		
1724	TGTGCTCACGACCGTGGTGTGATTTCTTATGATGGTGGAGATCTGAAGAGGAATTTGGA	1783
Db		
1796	CAACTGTATTCAATGTGTGAAAAATGAAATGTAGATTCGACAAAGGAATTTCTTTTCAGA	1855
Qy		
1784	TAACCTTATAGAAATTTGGTGGAAAAAGTGGGATCCTGATGGGAAGTGGGTACTCTGTCCAA	1843
Db		
1856	GCAATGTCGGATTTTATTTTTTATAGATTTAAAGATGCCAATCTGTGTGGATTGAGATGAAGC	1915
Qy		
1844	GGACGTTGGAATTTGATTTCTTTGCACTGCACAGCACAGTTTGTGAAATAGGAAGAGAGGC	1903
Db		
1916	TTTTTAAATGGCAAGCGCGAGTGAATTAACAGCCATGTTATTCAAACTCTGGTTGGAACTAAT	1975
Qy		
1904	TTTAGTATGGCAAGGACGAGTGTATGAGGAATGTATTCGATGGTGTGGTTGGCTCTGCT	1963
Db		
1976	GAATAGTATGTTTGAGAGAAGCTAATATPGACAAGAGANTGTTATGTGCCCACATTTAAATGA	2035
Qy		
1964	GAAGGTGATGAGAAAAAGGAAGCTGAATGGTTCGACAAATAAGGTAGTGCCTCAATGGGTGA	2023
Db		
2036	ATATATCGAAAAAGCTTACGTGTTCATTTGAGTCCCGATGTTGCAAGCCGGCTATTATTA	2095
Qy		
2024	ATATATGAAACAAGCCCATGATATCATTCGGGTTGGACACTAATATCCCTTCCAATGCTCTT	2083
Db		
2096	CTTTGTGGGGCCAAATTAATCAGAGGAGATTTGTTGAAAGCTCTGAATATCATTAATCTATT	2155
Qy		
2084	CTTTGTGGACCTTAACTCTCAGAGGAAATGATTTGGAAGCTGTGAATACAGAAAGTTATA	2143
Db		
2156	TAACTAATAGACACGCGAGGTGCACTTCTPAACGATATCCATAGCTTCAAGAGGGAAATT	2215
Qy		
2144	TAAGCTGATGAGCACTGCTGTCGCTTAAAGAATGATATTCGATCTTACGATAGAAATG	2203
Db		
2216	TAAAGGAGGCCAAATTAACCGCGTAGCATTTGCATTTAGTAAACGAGAAAGTGGGAAAGT	2275
Qy		
2204	CAGAAGAGGAAAGCTGAAATATCTGTCTCTGTGTGATGATGGCGGTGTTAATGTCAC	2263
Db		
2276	GGAAGACAGAGGTCTGTGGAGGAGATGATGATGATTAATAAAACAAGAGGAAGAAATTAAT	2335
Qy		
2264	CAGAAGAGGCCATTGAAGCAAATTAAGGGGAATTTTGAGAGGGCGGATAGAGAGCTGCT	2323
Db		
2336	GAAATTAATTTTGAAGAAATGCTAGCAATGTTTCTTAGAGCTTGTAAAGATGCATTTTG	2395
Qy		
2324	GGGGTTAGTTTTCAGGAGAGAACATCAAA---TTCCAAGAGCTTGTAAAGATTTGTTCTG	2380
Db		
2396	GAACATGTGTCAGTGTGTGAATTTTTTTTACGAAACGATGACGGGTTTACTGGAAACAC	2455
Qy		
2381	GAAATTTGATGCCATTGTGATCTATTTTACATGGAAGATGAUGGGTACACTTCCAATAG	2440
Db		
2456	GAATCTTGATCTGTGAAGGACATCATTTACAAACCGGTTGGTGGCTG	2502
Qy		
2441	GTTGTATGAACACTGTAAAGCCGATGTTTGAACAACCCATGATCTGG	2487
Db		

RESULT	6
CNU43904	
LOCUS	
DEFINITION	Cucurbita maxima ent-kaurene synthase B mRNA, complete cds.
ACCESSION	
U43904	
VERSION	
U43904.1	GI:1431869
KEYWORDS	
SOURCE	Cucurbita maxima (winter squash)
ORGANISM	
Cucurbita maxima	
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;	

Db 524 TGATCATATGACGAGGCCCTTGATTTATCAAGTCTAATATAGCTTCACTACTGATGA 583
QY 635 AAGTCAACCAATCTCCCAATGTTTTCACATCATATTTCTCTGTTTCTTGTAGTATCGGAA 594
Db 584 GAAACCAACGTTCTCCGGTGGATTTGACATTAATTTTCCCTGGCATGATTTGATATGCTAA 643
QY 695 AAACCTTGACATAAACCCTCTTTTCAAAACAACAGATTTTATGTTTGTATGCTACATAAGAG 754
Db 644 AGACTTGAATTTGAATCTACCTTGGCCACCGACGAACTGGATGCTTGGTTCGAAAGAA 703
QY 755 GGAATTTGAGCAAAA-----AGATGCCATTCAAATGAGATGATGATGATCTTGGCGTA 808
Db 704 AGAGTTGGAGCTGAGAGCTCGAGAGCAACTCTGAAGTGGAAAGCTTATTTAGCGTA 763
QY 809 TATCTCGAAGGATCGCGTAATTTATATGATTTGGAATATGTTGAAGAAATATCATGATGAA 868
Db 764 TGTTCAGAAGGAATGGAAAGTTTACAGACTGGGATATGCTCATGCAATATCAAGGAA 823
QY 869 AAATGGTTCTGTTTCACTCACCATCAGCAAGCTGCTGCTTTCAATTAATCATCAAAA 928
Db 824 GAATGGATCACTGTTTAAATTTCAATTTTCACTTTTGGACAGGTTTGGTAATGAGTCCCAAC 988
QY 929 TCTGTTGTTGTTTAAATTTTAAATTTCACTTTTGGACAGGTTTGGTAATGAGTCCCAAC 943
Db 884 TGAATGCTGTTTGAATTTTCTGCTTCACTTTCAAAAGTTTGGTGGCTCAGTTCCCAAC 943
QY 989 AGTTTATCTCATGATTTATTTATCCGACTTTCTATGTTGACACAATTTGAAGATTTAGG 1048
Db 944 AATATATCTCTTGATATATATGCTCGATTACACATGTTGATAGCTTCAAAATTCGG 1003
QY 1049 AATTTCAACCAATTCAGAGTGGAAATTTAAATTTTGAATGTTTGAATGAACATACAGATTTG 1108
Db 1004 AATTTGCTGGCAATTTCAAGAGGAGATTGAAGCGTATTAGATGAAACTTTACAGGTGTTG 1063
QY 1109 GGTGGAACAGATGACGAAATTTATTCAGGATTTGTAACATGCTTTAGTCCCTTTGCTT 1168
Db 1064 GATGCAAGGAGAGGAAATATATTTCTAGATGCTTCACTTTGCAATGCGCTTTTGGAAAT 1123
QY 1169 ATTAAAGATCAATGGGTATGAAGTTTCCCGAGATCCATTTGGCTGAAATTTACTAA----- 1222
Db 1124 GTTACGTTTGAAGGATATGATGTTCTTCAAGACCACTTCACTCAATTTTCAAGAGATAT 1183
QY 1223 -----TGAATTTAGCTTTGAAAGACGATATGAGCTTCACTCAATTTTCAAGAGATAT 1267
Db 1184 CTTTCCCAATGCTTTGGAGGATTTTAAAGACTTTGGGTGCTCGCTGAGTTATATAA 1243
QY 1268 TGGTGCATATATATATACAGAGGATTTATCTTCT---GGAAAAACAATCTTGAAGTC 1324
Db 1244 GGCCTCTCAGATTTATCAAGCAACCCGATGAACTCTGTTCTGGAATAATATAACTCTTGGAC 1303
QY 1325 AGCTGATTTCTCAAGAGATATATPCCACTGATTTCAAAACAGGCTTTCTAAA----- 1376
Db 1304 TAGTCGTTTCTTGAAGCATGGAATTTATCTAGTATTCAGTTTGTCTGTATGAAACCGATAG 1363
QY 1377 -TTAAATTCAGAGAGTGGAAAAATGCTCTTAAGTTCCTTCAATATACCGGTTTGAAGC 1435
Db 1364 TCTTGTAAACAGAGGCTGTTTAAAGCTCTTCAAGTTCCCTTATAATGCAACTCTGAAGC 1423
QY 1436 CATAAACACTAGACGAAATATACAGCTTTTACAAATGTAGACAATCAAGAAATCTGAABAC 1495
Db 1424 CCTAATAAGTAAAGGGCAATGGAAGTTTACAGTGGAGACATTTGAGGATTTCAAAATC 1483
QY 1496 TACATATCACTCAATATATAGTAAACATGATTAACCTTAAGTTGGCTTGTGGAAGATTT 1555
Db 1484 GCCATATGCTGCTTAAATTTTGGCCATCAAGATTTTCTGGAATTTGCTGTAGAGGATTT 1543
QY 1556 CTACACCTGCCAATCTATTTATCGTGAAGAAATTTAAAGTCTTTGAAAGTGGCTGTGAGA 1615
Db 1544 CAATACCTCTGAAACGATTTCACTTTAAAGAACTGGAAGAGCTTTCAAGATGGGTGTTGA 1603
QY 1616 GAATAGTTGACACAGCTCAAGTTTGTGAGCAAAAGACCGCTACTGTTATTTCTCTGT 1675
Db 1604 AAACAAATTTGACGAGTTGAAATTTTTCAGACTGCACCTAGGTTACTGTTTGTCTG 1663

QY 1676 TGCTCAACACATTTCTGCTCCGAATTTATCAGATCGGCTATTTTCATGGGCCCAAAATGG 1735
Db 1664 GGCAGCGACCTTACTGATCTCTGAATCTTCAATGATCTCGCATAGCATGGCACAATGG 1723
QY 1736 CATATTAATCACTAGTATGATGATCTTTTGTATATCGGTGGTACAATCGATGAATGAC 1795
Db 1724 TGTGCTCAGGACCGGTGGTGTATGATTTCTATGATGGTGGAGGATCTGAAGAGGAATTGA 1783
QY 1796 CAACCTGATTTCAATGTTGAAAAATGGAATGTAGATCTGCACAGAGGATTTGTTTTCAGA 1855
Db 1784 TAACTTTATAGATTTGTTGGTGAAGTGGATCTCTGATGGGGAAGTGGTTACTGTTTCCA 1843
QY 1856 GCATGTTCCGATTTTATTTTATAGCAATTAAGATGCAATCTGTTGGAATGGAGATGAAGC 1915
Db 1844 GGACGTTGAGATTTGATTTCTTGCATGTCACAGCACAGTCTTGTGAAATAGGAAGAGC 1903
QY 1916 TTTTAAATGGCAAGCGCGGATGTAATCTAGCCATGTTTATTAACCTTGGTTGGAACTAAT 1975
Db 1904 TTTAGTATGGCAAGACGAGTGTATGAGGAATGTTATCGATGGTTGGTTGGCTCTGCT 1963
QY 1976 GAATGATGTTGTGAGAGAGCTATATGCAACAAGAGATGCTTATGTCACCAATTAATGA 2035
Db 1964 GAAGGTGATGAGAAAGGAGCTGAATGCTGACAAATAAGAGTAGTGCATCAATGGGTGA 2023
QY 2036 ATATATCGAAACGCTTACGTTGATTTGATAGGCCCGGATTTCAAGCGGCTATTTTA 2095
Db 2024 ATATATGGAAACAAGCCCATGATCATTCGGTTTGGGACCTATAATCTTCCAATGCTCT 2083
QY 2096 CTTTGTGGGCCCAATTTATCAGAGGAGATTTGTAAGCTCTGAATATCATATCTATT 2155
Db 2084 CTTTGTGGACCTTAACTCTCAGAGGAATGATTTGGAAGCTGTGAATACCAGAGTTATA 2143
QY 2156 TAAGCTAATGAGCACGCGAGGTCGATCTTCAAAACGATATCCATAGCTTCAAGAGGAAAT 2215
Db 2144 TAAGCTGATGAGCACTGCTGCTGCTTAAAGATGATATTCGATCTTACGATGAGATG 2203
QY 2216 TAAGGAAGCAATTAACCGGCTAGCATTTGCAATTTGAGTAAACGAGAAAGTGGGAAAGT 2275
Db 2204 CAAAGGGAAGAGCTGAATATTTCTGCTCTGTGGATGATGATGGGCTGTTATGTCAC 2263
QY 2276 GGAAGGAAGAGTGTGGAGGAGATGATGATGATTAATAACAGAGGAAAGAAATTAAT 2335
Db 2264 CAAAGAGAGGCAATTTGAAGCAATTAAGGGATTTTGAAGGCGCATGAGAGAGTGT 2323
QY 2336 GAAATTAATTTTGAAGAAATTTGTTAGCATTTGCTTAGAGCTTTGTAAGATGCAATTTG 2395
Db 2324 GGGGTTAGTTTTCAGAGGAAACATACAA--TTCCAAGAGCTTTGAAGATTTGTTCTG 2380
QY 2396 GAACATGCTCAGCTGTTGAAATTTTTCGGAACAGATGACGGTTTACTGGAACAC 2455
Db 2381 GAAATGATGCTCAATTTGTAATCTATTTTACATGGAAGATGATGGGTACATCTCAATAG 2440
QY 2456 GATTCCTGATCTGTAAGGACATCATTTTCAACCGCTTGGTCTG 2502
Db 2441 GTTGTGACACTGTAAAGCCATGTTTGAACACCCATGGATCTG 2487

RESULT 7
AF034774
LOCUS
DEFINITION
Arabidopsis thaliana ent-kaurene synthase (GA2) mRNA, complete cds.
ACCESSION
AF034774
VERSION
AF034774.1
KEYWORDS
GI:3056724
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 2506)
Yamaguchi, S., Sun, T., Kawaide, H. and Kamiya, Y.
The GA2 locus of Arabidopsis thaliana encodes ent-kaurene synthase

of gibberellin biosynthesis
Plant Physiol. 116 (4), 1271-1278 (1998)
98205064
9536043
2 (bases 1 to 2506)
Yanaguchi S., Sun, T.-P., Kawaide, H. and Kamiya, Y.
Direct Submission
Submitted (14-NOV-1997) Botany, Duke University, Research Dr.,
Durham, NC 27708, USA
Location/Qualifiers
1..2506
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="1"
1..2506
/gene="GA2"
68..2425
/gene="GA2"
/function="terpene cyclase"
/codon_start=1
/product="ent-kaurene synthase"
/protein_id="AAC39443.1"
/db_xref="GI:3056725"
/translation="MSINLRSSGSSPISATLERGLDSEVQTRANNVSFEQTEKIRK
MKVEIUSVAYSATDSWAMVPSQNPAPFPQVKWLLDNOHEDGSLNDHDSHOSL
KLVLSSTLAIKWKWIGIERQINLGLOFIELNSALVTDETIQKPGFDIIPRGM
KYARDNLNLTIPLGSEVVDMDIRKRDLDKCDSEKFSKREAYLAVLEGTRNLKDWDL
IVKYORNGSLFDPATTAATAFTQPDNGDCLRLCSLQKFEAAVPSVYPPDVARLS
ILVLESGLDIRDKTEIKSLIDETRYWLRGDBEI CLDLATCALAFLLLAHGYDS
YDPKPAEBSGFSDTLEGVYKNTFSVLELFKAAQSPHESALKKQCWTQVYLEMEL
SSWKTYSVRDKLKEVEDALAFPSYASLERSDHRKILNGSAVENTRVKTSVRLHN
ICTSDIILKLVDFNFCQSIRHEBERLDRI VERNLQELKFAKOKLAYCVFSGAATL
FSPSLDARTISWAGGVLTTVDVDFEDVGGSKLENLHLVKEWDLNGVPEYSEHV
EIIIFSVLRDITILETGKAFYQGRNVTHIIVKWLIDLLKSMIREAESWSSDKSPSLSD
YMENAYISFALGPVILPATYILGIPLEKTVDSHQYNOLYKLVSTMGRLNDIOGPKR
ESAEKGNLAVSLHMKHEDNRKKEVIESMKGLAERKRBEELHLKLVLEKGSVVPRECK
EAFLLMSKVLNLFYRKDDGTDSNLSMLSVKSVIYEPVSLQKESLT"

Query Match 26.2%; Score 732.4; DB 8; Length 2506;
Best Local Similarity 59.5%; Pred. No. 2.5e-133;
Matches 1356; Conservative 0; Mismatches 876; Indels 48; Gaps 5;

QY 270 GGACAAACTAATCCCACTAATCTCATCATATATATACAAACCAAGAACGATCCAAACACAG 329
DB 143 GTACAGACAAGAGCTAACAAATGTCAGCTTTGAGCAACAAGAGAGAGATTAGGAAGATG 202
QY 330 TTTAAATAATGTAGAAATTTCTGTTTCTTATATGACACAGCATGGGTAGCCATGGTCCCT 389
DB 203 TTGGAGAAAGTGGAGCTTTCTGTTTCCGCTTACGATACCTAGTTGGTAGCAATGGTTCCA 262
QY 390 TCTCCAAACTCACCAATCCCTGTTTCTCCCTGAGTGTCTCAATTGGTTAAATAAAT 449
DB 263 TCACCGAGCTCCCAAAATGTCACCTTTTCCACAGTGTGTGAATGGTTATTGGATAAT 322
QY 450 CAGCTTAATGATGGTTTCATGGGCTTTGTTTAAATCACACTCATATATCAATATCACCGCTG 509
DB 323 CAACATGAAGATGGATCTTGGGAGCTTGATAACC-----ATGACCAATCAATCTCTT 373
QY 510 CTTAAAGATTTCTTAATCTCAACATTAAGCANGATTGTTGTCATTAATAAAGATGGAATGTT 569
DB 374 AAGAAGGANGTGTATCATCTACACTGGCTAGTATCTCGCGTTAAAGAAGTGGGAATTT 433
QY 570 GCGGAGAGATCAATAATAAAGTCTTAAGTTTATTGAGTCAAAATCTTGCTTCAGCTACT 629
DB 434 GGTGAAGACAAATAAACAAGGCTCCAGTTTATTGAGCTGAATTTCTGCAATTAGTCACT 493
QY 630 GAAAAAAGTCAACCATCTCCCATTTGGTTTATGACATCATATTTCTGTTTGGTTGAGTAT 689
DB 494 GATGAACCAATACAGAAACCAACAGGGTTGATATATATTTCTCTGGGATGATTAATAT 553
QY 690 GCGAAAAACTTGACATATAAACCTCTCTTCAAAACAAACAGATTTAGTTTGAATGCTACAT 749

ORIGIN

DB 554 GCTAGAGATTTGAATCTGACGATTCATTTGGGCTCAGAAAGTGGTGGATGACATGATACGA 613
QY 750 AAGAGGGAATTCGAGCAAAAAAG-----ATGCCATTCAAAATGAGATGATGATGATAC 800
DB 614 AAAAGAGATCTGATCTTAAATGTGATAGTGAAGATTTTCAAAGGGAAGAGAGATAT 673
QY 801 TTGGCGTATATCTCTGAAGGACTCGGTAAATTTATATGATGGAATATGATGGAAGAAATAT 860
DB 674 CTGGCCTATGTTTATAGAGGGGACAAAGAACCTTAAAGATTTGGGATTTGATAGTCAATAT 733
QY 861 CAGATGAAAAATGTTCTGTTTTCACCTCACCATCAGCAACAGCTGCTGCTTCATTAAT 920
DB 734 CAAAGGAAAAATGGGTCTACCTGTTGATTTCCAGCCACACACAGCAGCTGCTTTACTCAG 793
QY 921 CATCAAAATCCTGGTGTCTTAAATTAATTAATTTGGAACAAGTTTGGTAATGCA 980
DB 794 TTTGGGAATGATGTTGTTCTCCGTTATCTCTGTTCTCTCTTCAGAAATTCAGGCTGCA 853
QY 981 GTCCCAACAGTTTATCTCATGATTTATTTATTCGACTTTTCTATGGTTTGACACAAATGAA 1040
DB 854 GTTCTCTCAGTTTATCCATTTGATCAATATGCAAGCCCTTAGTATATAATTTGTCACCTCTTGA 913
QY 1041 AGATTAGGAATTTCAACCAATTTGAGTGGAAATTAATAATGTTTATAGTGAACATATAC 1100
DB 914 AGCTTAGGAATTTGATAGATTTTCAAAACCGAAATCAAAAGCATATTTGGATGAAACCTAT 973
QY 1101 AGATTTGGGTGGAAACGAGATGAGCAATAATTTATGATGTTTAAACATGCTTTAGCC 1160
DB 974 AGATATTGGCTTCTGGGGATGAGAATAATGTTTGGACTTGGCCACTTTGCTTTGGCT 1033
QY 1161 TTTCCGTTTATTAAGGATCAATGGGTATGAAGTTTCCCAGATCCATTTGGCTGAAATTTACT 1220
DB 1034 TTCCGATTTATGCTTGTCTCATGGGCTATGATGTTCTTACGATCCGTAAACCACTTTGCA 1093
QY 1221 AATCAATTAGCTTTGAAAGACGATATGCGACTCTTGAACATATCATCGCTCATATA 1280
DB 1094 GAAAGATCTGGTTTCTCTGATCTTTTGAAGGATATGTTGAAGATAAGTTTCTTCTGTGTTA 1153
QY 1281 TTATACCAAGAGGAT-----TTATCTTCTGGAACAAATCTTGAAGTTCAGCTGATTTCC 1335
DB 1154 GAATTTATTTAAGGCTGCTCAAGATTATCCACATGAATCAGCTTTTGAAGAGCAGTGTTGT 1213
QY 1336 TCAAGAGATATATTCACCTGATTCAA-----ACAGGCTTTCT 1373
DB 1214 TGGACTAAAACATATCTGGAGATGGAATTTCTCCAGCTGGGTAAAGACCTCTGTTCCGAGAT 1273
QY 1374 AAATTAATTTCAAAAGAGTGGAAATGCTCTTAAAGTTCCCTATCAATACCGGTTTGA 1433
DB 1274 AAATACCTCAAGAAAGAGTTCGAGGATGCTTCTTGTCTTCCCTCTATGCAAGCCTAGAA 1333
QY 1434 CGCATAAACACTAGACGAATAATATAC---AGCTTTTCAATGTAGACATACAAAGATTTCTG 1490
DB 1334 AGATCAGATCACAGGAGAAAAATACTCAATGGTTCTGCTGTGGAACACACAGATTACA 1393
QY 1491 AAAACTCATATCATCATCAATAATTTAGTAAACATGATTAACCTTAAGTTGGCTGTTGAA 1550
DB 1394 AAAACCTCATATCGTTTGCACATATTTGCACCTCTGATATCTCTGAAGTTAGCTGGAT 1453
QY 1551 GATTTCTACACCTGCGCAATCTATTTATCGTGAAGATTAAGGCTCTTGAAGCGTGGGT 1610
DB 1454 GACTTCAATTTCTGCGAGTCCATACACCGTGAAGAAATGGAACGTCTTGAATAGTGGATT 1513
QY 1611 GTAGAGAAATAGTTGACCGAGCTCAAGTTTGTAGGCAAAAGACCGCTACTGTTATTTTC 1670
DB 1514 GTGAGAAATAGATTGACGAACTGAATTTGCCAGACAGAGCTGGCTTACTGTTATTTTC 1573
QY 1671 TCTGTTGCTGCAACATTTCTGCTCCCGAATTAATCAGATCCGCGTATTTTCATGGGCCAAA 1730
DB 1574 TCTGGGGCTGCAACTTATTTTCTCCAGAACTATCTGATGCTCGTATATCGTGGGCCAAA 1633
QY 1731 AATGGCATATTAACTACAGTACTTGTGATGACTTTTGTGATATCGTGGTGAATTCGATGAA 1790

Db	1634	GGTGGAGTACTTACAAACGGTTGTAGACCACTTCTTTGATGTTGGAGGTCCTCAAGAGAA	1693	Upadhaya,N.M. Direct Submission Submitted (21-JUL-2003) Rice Functional Genomics Group, Genomics and Plant Development Program, CSIRO Plant Industry, Chr. Barry Drive and Clunies Ross Street, GPO Box 1600, Canberra, ACT 2601, Australia
Qy	1791	TTGACCAACCTGATTCATTAATGTTGAAATAATGGAATGATGATGTCGACAGGATGTTCT	1850	
Db	1694	CTGGAACACCTTACATACACTTGGTCGAAAGTGGGATTTGAACGGTGTTCCTCGAGTACAGC	1753	
Qy	1851	TCAGAGCATGTTGGATTTATTTTATTTAGCATTAATAAGATGCAATCTGTTGGATTTGGAGAT	1910	
Db	1754	TCAGAACATGTTGAGATCATATTTCTCAGTTCTTAGGGGACACCATCTCGAAACAGGAGAC	1813	
Qy	1911	GAAGCTTTTAAATGCGACGCGGATGCTAATAGCCATGTTTATTCAACTTGGTTGGAA	1970	
Db	1814	AAAGCATTTCACTATCAAGAGCGCAATGTGACACACCAATTTGTGAAATTTGGTTGGAT	1873	
Qy	1971	CTAATGAATAGTATGTTGAGAGAACTATATGGAACAAGATGCTTATGTGCGCAATTA	2030	
Db	1874	CTGCTCAAGTCTATGTTGAGAGAAACCGGATGCTCAGTGAAGTCAACCAACGAGCTTG	1933	
Qy	2031	AATGAATATATGGAACCGCTTACCTGTCATTTGATAGGCCGCTTGTCAAGCCGCT	2090	
Db	1934	GAGGATTCATGGAATAACCGTACATATCATTTGATAGGCCAATTTGTCTCCAGCT	1993	
Qy	2091	ATTTACTTTTGTGGGCCCCAAATATCAGAGGAGATTTGTTGAAGCTCTGAATATCATAT	2150	
Db	1994	ACCTATCTGATCGGACCTCCACTTCCAGAGAGACAGTGTGATGACCAATATAATCAG	2053	
Qy	2151	CTATTAAAGTAAATGAGCAGCGGTCGACTTCTTAACGATATCTAGCTTCAAGG	2210	
Db	2054	CTCTCAAGCTCGTGAGCACTATGGTCTCTTCTTAATGATCATCAAGGTTTAAAGA	2113	
Qy	2211	GAATTTAGGAAGGCAATTAACCGGCTAGCATTTGATTAACCGGAGAAAGTGGG	2270	
Db	2114	GAAGCGCGGAAGGAGAGCTGAATGCGGTTTCTGCAATGAACACGAGAGACAT	2173	
Qy	2271	AAAGTGAAGAGAGAGGTTCTGAGGAGATCATGATGATTAATAACAGAGAGAA	2330	
Db	2174	CGCAGCAAGAGATGATCATAGATCGATGAAAGTTTACGACAGAGAGAAAGGAGAA	2233	
Qy	2331	TAAATGAATTAATTTTGAAGAAATGTTAGCAATTTGTTCTTAGAGCTTTGAAGATCA	2390	
Db	2234	TTGCAATAGCTAGTTTGGAGGAGAAAGGAGTGTGTTTCCAGGGAATGCAAGAGCG	2293	
Qy	2391	TTTGTGAACATGTCACGTTGTTGAATTTTTCAGCAACCAATGATGCGGTTTCTGGA	2450	
Db	2294	TTCTTGAATGAGCAAGTGTGATTTTATTTTACAGGAGACGATGATTCATCA	2353	
Qy	2451	AACACGATTTGATGATGTTGAAGGACATCATTTTCAACCGGTTGTTGATGAATGAA	2510	
Db	2354	AATGATCTGATGATGTTGTTTAAATCAGTGTACGAGGCTGTAGCTTACAGAGAA	2413	
RESULT 8	AY347876	2271 bp	linear	PLN 05-AUG-2004
LOCUS	AY347876			
DEFINITION	Oryza sativa (japonica cultivar-group) map BAC OSJNBa0070C17			
ACCESSION	AY347876			
VERSION	AY347876.1			
KEYWORDS	ent-kaurene synthase 1A (OsKSI1A) mRNA, complete cds.			
SOURCE	Oryza sativa (japonica cultivar-group)			
ORGANISM	Oryza sativa (japonica cultivar-group)			
REFERENCE	1 (bases 1 to 2271)			
AUTHORS	Upadhaya,N.M.			
TITLE	Isolation and characterization of a Ds-tagged rice (Oryza sativa L.) GA-responsive dwarf mutant defective in an early step of the gibberellin biosynthesis pathway			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 2271)			
AUTHORS	Margis-Ponheiro,M., Zhou,X.-R., Zhu,Q.-H., Dennis,B.S. and			

misc_difference	1087	gene="OsKSI1A"	/note="compared to genomic sequence in GenBank Accession
		Number AL731610"	
		/replace="t"	
ORIGIN			
Query Match	20.1%;	Score 559.8;	DB 8; Length 2271;
Best Local Similarity	56.0%;	Pred. No. 1.7e-99;	
Matches 1215;	Conservative	0; Mismatches 902;	Indels 54; Gaps 6;
Qy	308	CAAAGAAGGATGCTCAAAAACAGTTTAAATAATGTAGAAAATTTCTGTTTCTTTCATATGACAC	367
Db	21	CCAGGCAAGGACAAAGGAGCCAGCTCCAGACATTTGGAGTTATCTACGCTTTTGTACGACAC	80
Qy	368	AGATGGGTAGCATGCTGCTTCTTCAAACTCACCACATCGCTGTTTCTTCTGATG	427
Db	81	GGCGTGGTGGCCATGCTGCCACTGCGGGGTTCCTCCGTCAGCATCCGCTTCCCAATG	140
Qy	428	TCCTCAATGCTTAAATAATTAATCAGCTTAAATGATGTTTATGTTGTTTCTTCTTAAATCACAC	487
Db	141	CGTGGAGTGAATATTACAAACACGAGGATGATGATCTTGGGT-----AC	188
Qy	488	TCATAATCAATATCACCCTGCTTAAAGATTTCTTATCTTCAACATAGCATGATTTGT	547
Db	189	AAGAGGATTCGGCGTGGCGGTCAAGAGATGTTCTGCTCTACGTTGCGTGTGTCT	248
Qy	548	TGATTTAAAGATGGAATGTTGGGAGAGATCAATAATAAATAAGGTTCTAAGTTTATTTGA	607
Db	249	TGCATCAAGAGATGGAATGTTGGGAGGACATCAGGAGGAGCATGATTTATTTG	308
Qy	608	GTCAATCTTCTGCTTACGCTACTGAAATAAGTCAACCATCTCCATTTGGTTTGAATCAT	667
Db	309	AAGGAATTTCTCCATTTCCATGGATGAGCAGATTTGCTGCTCTGTTTCAACATCAC	368

QY 668 ATTTCCTGTTTGGTTTGAAGTATGCGAAAAAATTGGACATATAACCTCCCTTTCACAAACAAC 727
Db |||||
QY 369 TTTTCTGTTATGCTTATGCTTGGATGGGATGATTTAGATTTCCCTGTGACAGAC 428
Db |||||
QY 728 AGATTTTATGTTGATGCTACATATAGAGGAAATTGAGCAAAAAA-----GATGCCA 778
Db |||||
QY 429 GGAATGTTGATAGATTTCTTCCCTCCGGAGATAGAACTGGAAAGAGAGGCTGGAGACCA 488
Db |||||
QY 779 TTCAATGAGATGATGATGATGCTTGGCTATATCTCTGAAGACTCGTAAATTTATATGA 838
Db |||||
QY 489 TTTCTATGGAAGAAAGGCAATATATGCTTATGTCACAGAGGATTAGGAAACCTGTTGGA 548
Db |||||
QY 839 TTGGAATATGTTGGAAGAAATATCATGATGAAAAATGGTTCTGTTTCAACTCACATCAGC 898
Db |||||
QY 549 GTGGGATGAGATTTATGATGTTTCCAGAGGAAGATGGAATTTTCAACTGCTCTCCAC 608
Db |||||
QY 899 AACAGCTGCTCTTTCATTAATCATCAAAAAATCCTGGTTGCTTATATTTAAATCACT 958
Db |||||
QY 609 AACTGCTGCCACATTAGTCAACCACTATAACGATAAAGCCCTCCAGTACTTAAATTTGCT 668
Db |||||
QY 959 TTTGGCAAGTTTGGTATGATGAGTCCCAACAGTTTATCCTCATGATTTATTTATCCGACT 1018
Db |||||
QY 669 GGTAGTAAATTTGGCAGTGAGTACCAACAGTGTATCCATTAACATATATTTGTCCAGCT 728
Db |||||
QY 1019 TTTCTATGTTGACCAATTTGAAGATTTAGGAATTTTCAACCAATTTTCAAGTGGAAATTTAA 1078
Db |||||
QY 729 TTTATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 788
Db |||||
QY 1079 AAATGTTTATGATGAACATACATGTTGGTGGAGAGGATGAGCAATATTTATCATGGA 1138
Db |||||
QY 789 GAGATCCTAGACACAACTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 848
Db |||||
QY 1139 TGTGTTAACTATGCTTTAGCTTTTGGTGTATTAAGGATCAATGGGTATGAGTTTCCCC 1198
Db |||||
QY 849 CATTAACACATGTCATGCTTTTCCCTGTTTGAAGATGATGATGATGATGATGATGATGAT 908
Db |||||
QY 1199 AGATCCATTTGGCTGAATTTACTAA-----TGAATTAGCTTT 1234
Db |||||
QY 909 AGTCGAGTTATCTCTGCTGCTGAGCTTCCAGTTTCCGTTGAATCACTTCAAGATATTT 968
Db |||||
QY 1235 GAAAGCAATATGACGCTCTTGAACATATCATGCTGCTCAATATATATACCAAGAGGA 1294
Db |||||
QY 969 AAATGATTAAGAAATCTCTAAATCGAAATTTATACAGGCTTCAAAAGTCAAGTAAATCAGAAA 1028
Db |||||
QY 1295 TTTATCTTCTGGAACAAATCTTGAAGTCAAGTCTGATTTCTCTCAAGAGATATATCCAC 1354
Db |||||
QY 1029 TGAATCAATCTTAGATAGCATAGCTCTTGGTCAAGTATGCTTTATTTGAAGAAAGTGTGAG 1088
Db |||||
QY 1355 TGATTCAAAAGGCTTTCTTAAATTTAAATTTCAAGAGGTTGGAATAATGCTCTTAAGTTCCC 1414
Db |||||
QY 1089 CTCTAATGGGTTGAAAAAAGCTCCATCTTTGAAGAGATGATGCTCTTAAGTTTCC 1148
Db |||||
QY 1415 TATCAATACCGGTTTGAAGCGATATACACTAGACGAAATATACAGCTTTACAAATGAGA 1474
Db |||||
QY 1149 CTTTCAACCACTGGACCGCTTAGATACACAGAGGAAATTTGAAGTTTGTGATGAAA 1208
Db |||||
QY 1475 CAATACAAGAAATTTCTGAAAAATCATATCACTCATCAAAATTTAGTAACACTGATTAACCT 1534
Db |||||
QY 1209 GGATTTCCAGATTTAAAGCGAAT---ACTTGTCTCTCATGCCAATCAAGATATTTCT 1265
Db |||||
QY 1535 AAGTTGGCTGTTGAAGATTTTACACCTGCCAATCTATTATTCGTCAAGAAATTTAAAGG 1594
Db |||||
QY 1266 AGCTTTGGCTGTTGAAGATTTTCAAGTATGTTCTCAATCTATATACAGGATGAACTTAATTA 1325
Db |||||
QY 1595 TCTTTGAAGGTTGGTGTAGAGATTAAGTTGGACAGCTCAAGTTTGTCTAGGCAAAAGAC 1654
Db |||||
QY 1326 TCTTGAAGTGTGGTGAAGATGAAAGAGTTCGATCAGCTGCCATTCGACGCCGAGAGTT 1385
Db |||||
QY 1655 CGCTACTGTTATTTCTCTGCTGCTCAACACATTTTCTGCTCCCGAATTTATCATAGCGGG 1714
Db |||||
QY 1386 GACATATTTGCTACTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1445
Db |||||

QY 1715 TATTTTCATGGCCAAAATGGCATATTAACCTACAGTAGTGTGATGACTTTTGTGATATCGG 1774
Db |||||
QY 1446 CATTTGCATGGGCTAAAAATGGTGTACTGACCAACTGTTGTTGATGACTTTCTTTGATCTTGG 1505
Db |||||
QY 1775 TGTACAAATCGATGAATTTGACCAACCTGATTCATTAATGCTTTGAAAAATCGAATGTAGATCT 1834
Db |||||
QY 1506 GGGATCAAAAGAGAACTAGAAACCTCATTTGCTTTAGTGAAGAGUGGATGGACAT-- 1563
Db |||||
QY 1835 CGAACAGATTTGTTGTTGTCAGAGCATGTTTCGGAATTTTATTTTATAGCAATTAAGATGCAAT 1894
Db |||||
QY 1564 -CAAGAGAGGTTTCTACTCAGAACAAAGTAAGATAAGTTTCTGCTATTTATCTACAGT 1622
Db |||||
QY 1895 CTGTTGATTTGAGATGAGCTTTTAAATGCGACGCGCATGATTAATAGCCATGTTAT 1954
Db |||||
QY 1623 GAACAGCTTTGGAGCAAAAGGCTTCTGCAATTAAGAGGCGTGTATGTACAAAACCTTAAC 1682
Db |||||
QY 1955 TCAAACTTGGTTGGAACCTAATGAATAGTATGTTGAGAGAACTATATGACAAAGAGATGC 2014
Db |||||
QY 1683 AGAAATATGTTTATGTTCTGATGAGGTCTATGATGCCGAAGCTGATGGCAGAGACAA 1742
Db |||||
QY 2015 TTATGTCGCAACATTAATGAATATATGGAAGAAACGTTTACCTGTCATTTGCAATTTAGGCC 2074
Db |||||
QY 1743 ATATGTCGCGCAATGGAAGAAATATATGGCAATATGCTGTTGCTCTCATTTGCACTGGGACC 1802
Db |||||
QY 2075 GATTTCAAGCCGCTATTTTACTTTTGGGGCCCAATTTATCAGAGAGATTTGTTGAAG 2134
Db |||||
QY 1803 TATTTGCTCTCCGACTCTGATTTTCGTAGGACCGAAGCTTCAAGAGGATGTCGTCAAGGA 1862
Db |||||
QY 2135 CTCGAAATATCATATCTATTTTAACTGATGACGACGAGGTCGACTTCTTAAACGATAT 2194
Db |||||
QY 1863 TCATGAGTCAATGAAATTTATTTAGCTGATGAGCACTTGTGGGCTCTCTGATGATGACAG 1922
Db |||||
QY 2195 CCATGCTTCAAGAGGGAATTTAGGAAGCAATTAACCGGTAGCATTTGCAATTTGAG 2254
Db |||||
QY 1923 CCAAGGCTTTGAGAGGGAGAGCCCTGGAGGGAAGCTGACAGTGTCTCACTGCTTTGTTCA 1982
Db |||||
QY 2255 TAACGAGAAAGTGGGAAAGTGGGAAGAGGTTGTGGAGAGATGATGATGATGATGATGAT 2314
Db |||||
QY 1983 TCACAGTGTGTTTCTATCTCTCAGAGGCTTAAATGAAAGCCAGAAATCCATAGA 2042
Db |||||
QY 2315 AAACAAGAGAAAGAAATTAATGAAATTAATTTTGAAGAAATGTTAGCATTTGTTCTAG 2374
Db |||||
QY 2043 CACTTCCAGGAGGAACCTTGTAGATTTGCTTGTGGAGACAGGT---CCTGTTCTCTAG 2099
Db |||||
QY 2375 AGCTTTGAAGATGCAATTTTGGAAACAGTGTACGTTTGAATTTTACGCAACGA 2434
Db |||||
QY 2100 GCCATCAAGCAGCTGTTCTGGAAGATGTCGAAGATTTGTTCAATGTTCTACTCTCGGAC 2159
Db |||||
QY 2435 TGACGGGTTTA 2445
Db |||||
QY 2160 TGATGGGTTTA 2170
Db |||||

RESULT 9
AY347877
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) clone OsKS1B mutant clone 1
ent-kaurene synthase 1B (OsKS1B) mRNA, complete cds.
ACCESSION
AY347877
VERSION
AY347877.1 GI:34452215
KEYWORDS
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Bhrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 2283)
AUTHORS
Margis-Ponheiro, M., Zhou, X.-R., Zhu, Q.-H., Dennis, E.S. and
Upadhyaya, N.M.
Isolation and characterization of a Ds-tagged rice (Oryza sativa
L.) GA-responsive dwarf mutant defective in an early step of the
gibberellin biosynthesis pathway
Unpublished

1957 ATCTTTGAAGAGTGAAGTATGCTCTTAAGTTTCCCTTCTACACCACTGGACCGTCTA 1916
 1440 AACACTAGAGAAATATACAGCTTTTCAATATGACAAATACAGAAATCTGAAACTACA 1499
 1917 GATCACAGAGGAAACATTTGAACGTTTGTATGCAAGAGATCCAGATGTTAAAGACGGAA 1976
 1500 TATCACTCATCAATATATAGTAACTGATTAACCTAAGTTGGCTGTTTGAAGATTTCTAC 1559
 1977 T---ACTTGTCTCTCATGCCAATCAAGATATCTAGCTTTGGCTGTTTGAAGATTTCTAGT 2033
 1560 ACCTGCCAATCTATTTATCGTGAAGAAATTAAGAGTCTTGAAGTGGTGGTGTAGAGAAAT 1619
 2034 AGTTCTCAATCTATATACAGAGTGAATTAATATCTTGTAGTGTGGTGAAGATCAA 2093
 1620 AAGTTGGACAGCTCAAGTTTGTAGGCAAAAGACCGCTACTGTTATTTCTCTGTTGCT 1679
 2094 AAGCTCGATCAGCTGCCATTTGCACGCCAGAGTTGACATATGCTACTTTCTGCTGCT 2153
 1680 GCAACACTTTGCTCTCCCGAATATACAGATGCGCTATTTTCATGGCCCAAAAATGGCATA 1739
 2154 GCTACCATATTTCCCGGTAATGTTCTGAAGCCGCATTCGATGGCTTAAATATGCTGTA 2213
 1740 TTAACACAGTAGTGTGACATTTTGTATATCGGTGTGATCAATCGATGAATGTGACCAAC 1799
 2214 CTGCAACTGTTGTTGATGACTTTCTTGTATCTTGGGGATCAAAAGAGAACTAGAAAAC 2273
 1800 CTGATCAATGTTGTTGAAATGGAATGTAGATGTGCAAGAGGATGTTGTTACAGCAT 1859
 2274 CTCAATGTTTGTAGTTGAGAGTGGATGGACAT---CAAGAGGAGTTCTACTCAGAACAA 2330
 1860 GTTCGGATTTATTTTGTAGCTATAAAGATGCAATCTGTTGATTTGGAGATGAAGCTTTT 1919
 2331 GTAAAGATGTTTCTGCTATTTATCTACAGTGAACCACTTGGACCAAGGCTTCT 2390
 1920 AATGGCAAGCGCGGATGTAATAGCCAGTATTTCAAACTGTTGTTGGAATCAATGAAAT 1979
 2391 GCATTTACAGGCGGTGATGTACAAAACCTTAACAGAAATATGTTATGCTGATGAGG 2450
 1980 AGTATGTTGAGAGAGCTATATAGGCAAGAGATGCTTATGTCGCACTTAATTAATGAAT 2039
 2451 TCTATGATCAGCGAAGCTGAATGGCAGAGGACAAAATATGTGCCGACAAATGGAAGAAAT 2510
 2040 ATGGAAAACGCTTACGTGTCATTTGCAATTAGGCGCGGATGTCAAAGCGCGCTATTTACTTT 2099
 2511 ATGGCAAAATGCTGTTGCTCAATTTGCTGAGGACCTATTTGCTCCGACTCTGATTTTC 2570
 2100 GTGGGCGCCAAATATCAGAGGAGATTTGTTGAAGCTCTGAATATCAATATCTATTTAAG 2159
 2571 GTAGGACCGAAGCTCCAGAGGATGTCGTCAAGGATCATGATCAATGAATTTATTTAGA 2630
 2160 CTAATGAGCAGCGAGGTGCACTTTCTAAAGATATCCATAGCTTCAAGAGGGAATTTAAG 2219
 2631 CTGATGAGCACTTTGTGGCGCTCTCCCTGAATGACAGCAAGGCTTTGAGAGGGAGACCTTG 2690
 2220 GAAGGCAAAATTAACCGCGTAGCATTTGATTAACGGAATAAGGAAAGTGGAAAGTGGAA 2279
 2691 GAGGAAAGCTGAACAGTGTCTCACTGCTTGTTCATCAGAGTGTGTTCTATCTCCATA 2750
 2280 GAAGAGGTTGTGGAGGAGATGATGATGATTAATTAATAAACAAGAGGAAGAAATTAAGAA 2339
 2751 GACGAGGCTAAATGAATGAAGCCGAAATCCATAGACACTTCCAGAGGAGAACTTTGTTAAGA 2810
 2340 TTAATTTTGAAGAAATGTAGCATTTGTTCTTAGAGCTTGTAAAGATGCATTTTGAAC 2399
 2811 TTGGTCTCTGGAAACAGGGT---GCTGTTCTTAGGCCATGCAAGCAGCTGTTCTGGAAG 2867
 2400 ATGTGTCACGTGTTGAATTTTTTTTAAAGCAAAAGATGACGGGTTTA 2445
 2868 ATGTGCAAGATTTGTCACATGTTCTACTCTCGGACTGATGGGTTTA 2913

RESULT 12
 AY347880

LOCUS AY347880 2529 bp mRNA linear PLN 05-AUG-2004
 DEFINITION Oryza sativa (japonica cultivar-group) map BAC OSUNB00137; BAC OSUNBA0052P16 ent-kaurene synthase like-4 (OsKS4) mRNA, complete cds.
 ACCESSION AY347880
 VERSION AY347880.1 GI:34452221
 KEYWORDS Oryza sativa (japonica cultivar-group)
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
 REFERENCE 1 (Bases 1 to 2529)
 AUTHORS Margis-Ponheiro,M., Zhou,X.-R., Zhu,Q.-H., Dennis,E.S. and Upadhyaya,N.M.
 TITLE Isolation and characterization of a Ds-tagged rice (Oryza sativa L.) GA-responsive dwarf mutant defective in an early step of the gibberellin biosynthesis pathway
 JOURNAL Unpublished
 REFERENCE 2 (Bases 1 to 2529)
 AUTHORS Margis-Ponheiro,M., Zhou,X.-R., Zhu,Q.-H., Dennis,E.S. and Upadhyaya,N.M.
 TITLE Direct Submission
 JOURNAL Submitted (21-JUL-2003) Rice Functional Genomics Group, Genomics and Plant Development Program, CSIRO Plant Industry, Chr. Barry Drive and Clunies Ross Street, GPO Box 1600, Canberra, ACT 2601, Australia
 FEATURES
 source
 Location/Qualifiers
 1..2529
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="4"
 /map="BAC OSUNB00137; BAC OSUNBA0052P16"
 1..2529
 /gene="OsKS4"
 1..2529
 /gene="OsKS4"
 /note="has an extra exon compared to OSUNB0052P16.10 in GenBank Accession Number AL62936.2; differs from OsKSIA sequence in GenBank Accession Number AY347876"
 /codon_start=1
 /product="ent-kaurene synthase like-4"
 /protein_id="AAQ72563.1"
 /db_xref="GI:34452222"
 /translation="WASPEAVARSLVAPRRRALGLPAAAAAPFVLCRRHN GMRPHYSFACSAELDTCEROLPGTGTAVMSQGVYVGRMVGENTSQINNGWEAR IURHLENPEFLFSSTDIAWAVPLPGTDHLQAPCEPCVEMWLNQHSNGSWNEF DSSAKDILLSTLACTIALLEKNVSGEQIRRGHTIAKNFSIVDDQIAAPIGFNLTF PAVNLAIKMGLEFFASEISDQILHRLMELKRIAGDESLGREATFATIAEGLESM VDWSEKFPQGNKSLFSPAAALVHYDDKALGYLSVNVKFGGVPVTVPLNI FSQLMVDPLVNIIGISRHFSIDIKRILDKTYLWSORDEEVLMDLPTCAMAFELRMN GYGVSSDDLSHVAEASTPHNSVGVLDLTKLLELYKASKVLSENEPILEKMGWSG SLKKEKLCDDIRGTFPILEVEYALKEFPYATLEPLDHKNINENFARAYOKIKTKNM PCHVNEDLALAAEDFPCCQSTYQNEIQHLESWEKENKLDQLEFFRKNLINSVLSAA TTPYELSDARIACASIALTAVDDFDVSGSKBQENLISLVEKWDQTHKVFISE NTKVAFYALSTVNLQGANASAVQNRDVTKNVSWLDLRSLATDAWQRSQPTPM ERYMKNISVTFALGPTILIALYFMGNLMEDIYKNAEYDELFLMNTCGRLONDIQSF ERECKDGLNLSVLLVDSKDVMSVEEAKEAINESISSORRELLRLVREDGVIKPSK KEMFNLYKTSIHVFSQADGFSFSPKEMNGAMNGVIFPEPLKTRGN"
 ORIGIN
 Query Match 17.2%; Score 479.6; DB 8; Length 2529;
 Best Local Similarity 53.8%; Pred. No. 8.7e-84;
 Matches 1185; Conservative 0; Mismatches 959; Indels 60; Gaps 7;
 285 ACTAATCTGATCTGATGATACCAACCAAGACGATCCAAACACAGTTAAATGTAGAA 344
 274 ACAAGTCAGATAAACATGGATGGGCTAGGATAGTACTTTAGGCACCTTGGAGAACCCGAG 333
 345 ATTCTCTGTTCTTTCATATGACACAGATGGGTAGCCATGTCCTCTTCTCAAACTCA--- 401

Db 334 TTCTTACCATCTTCATATGACATAGCATGGGTGGCTATGGTCCATTTGCCGGGCACTGAT 393
QY 402 ---CCCAATGCCCTTGTTCCTGAGTCTCAATTTGGTTAAATTAATCAGCTTAAT 458
Db 394 CATCTTCAAGCTCCAGTCTTCCTGAATGTGGAAATGATGCTACAAACCAACACAGT 453
QY 459 GATGTTCAATGGGTCCTTTTAATCACACATCATATAATCATATACACCCGTTGCTTAAAGAT 518
Db 454 AATGGGTGGTGGGTGTCA-----ATGAATTTGACTCATCCAGCAGCAGAT 501
QY 519 TCTCTATCTTCAACATTAAGCATGATGTTGTCATTAATAAAGATGAATGTTGGGAAAGAT 578
Db 502 ATTCTCCTATCCACTTTGGCATGTATTAATGACATGAGAAATGAATGTGGTTCGGAG 561
QY 579 CAAATAAATAAAGTCTAAGTTTATTCAGTCAAAATCTTGCTTCAAGTACTGMAAAAAGT 638
Db 562 CAAATAAGGAGGATTAATTTATCGAAGAATTTCTCCATTTGTTATGATGACCAG 621
QY 639 CAACCATCTCCATTTGACATCATATATTCCTGGTTTGTTCAGTATGCGAAAAAC 698
Db 622 ATTTGTGACCTATAGGCTTCAACCTCACAATTCCTGTCTATGTTAACTTTGCCATTAAG 681
QY 699 TTGACATATAACCTCTTTCAAAACAACACAGATTTAGTTTGTATGCTCATATAAGAGGAA 758
Db 682 ATGGTTTGGAAATTTCTTGGCAGTGAATTTAGTATGATCAGATTTCTTCACTCCGTGAT 741
QY 759 TTGAGCAAAAAAGA-----TGCCATTCAAATGAGATGGATGATATCTTGGCGTAT 809
Db 742 ATGGAAATGAAGACTGGCTGGTGAAGAACTTTTGGGGAAGAGGCATATTTCCGCTAT 801
QY 810 ATCTCTGAAGACT---CGGTAAATTTATGATTTGAATTTGGTGAAGAAATATCAGATG 866
Db 802 ATTTCTGAAGCTCAGAGAAGCATGGTGAATTTGGAGTGAATTAAGAATTCAGGGG 861
QY 867 AAAAAATGTTCTGTTTCACTCACCACACAGCTGCTGCTCTTCAATTAATCAACAA 926
Db 862 AAGAAATGATCATTTGTTCACTCCCGCTGCAACTGCTGCTGATAGTCCACAGATAC 921
QY 927 AATCCTGGTTGTTCTTAATTTAAATTTCACTTTTGGACAAGTTTGGTAAATGCAATTA 1046
Db 922 GATGATAAGCCCTGGGATACCTATATTTCTGCTGCTCAATAAATTTGGAGGTGAAGTACCA 981
QY 987 ACAGTTTATCTCATGATTTATTTATCCGATTTCTATGTTGTCACAAATTTGAAGATTA 1046
Db 982 ACCGTGATCCGTAAATATATTTCTCAGCTTTCAATGGTGGATCTCTGCTCAATTT 1041
QY 1047 GGAATTTCAACCAATTTTCAAGTGGAAATTAATAATGTTTGAATGAACATACAGATGT 1106
Db 1042 GGAATATCTCGGCACTTTCTAGTGATATAAGGCAATTTTGGATAAAGACATACATTTTA 1101
QY 1107 TGGGTGGAACGAGATGAGCAAAATATTTCATGATGTTGTAACATGCTTTAGCCTTTCCG 1166
Db 1102 TGGTCACAGAGAGATGAGGAAGTAATCTGGAATTTACCAACATGCGCAATTTGGCATTCGC 1161
QY 1167 TTATTAAGATCAATGGGTATGAAGTTTCCCGATCCATTTGGCTGAAATTTACTAATGAA 1226
Db 1162 CTTTTCGGTATGAACCGATATGTTGTTCTCTCAGATGACTTGTCCCATGTTGCTGAAGCC 1221
QY 1227 TTAGCTTTGAAACGCAATATGCG-----CTCTTGAACA 1262
Db 1222 TCACTTTTCACTAATCAGTTGAAGGATACCTTAGATGATACAAATCTTTATTAGAATTG 1281
QY 1263 TATCATCGCTCACATATATTATACCAAGAGGATTTATCTTCTGGAACCAAACTTTGAAG 1322
Db 1282 TACAAAGCTTCAAAAGTCAGTTTATCAGAAAATGAGCCAACTCTAGAGAAAATGGGTGTC 1341
QY 1323 TCAGCTGAATTTCTCAAGAGATAAATATCCATGATTCAAACAGGCTTTCTAATAATTAAT 1382
Db 1342 TGGTCAGGTAGCTTAATTTGAAGAAAAAATTTGTCTCCGATGACATCCGAGGAACACCAATC 1401
QY 1383 CACAAAGAGGTGGAATGCTTCTTAGTTCCCTATCAATACCGGTTTAGAACCCATAAC 1442
Db 1402 CTTGAGAGGTAGAAATATGCTCTCAATTTCCATTTTATGCCACGCTGGAACTCTAGAC 1461

QY 1443 ACTAGACGAATPATACAGCTTTCAATGTAGACAATAACAAGAAATTTCTGAAACTTACATAT 1502
Db 1462 CACAAGTGGAAATTTGAAATTTTGTATCCAGGGCTTATCAGAAGATAAAGACCAAAAC 1521
QY 1503 CACTCATCAATATATAGTAAACACTGATTTACCTTAAGTGTGGCTTGGCTTGAAGATTTTACACC 1562
Db 1522 ATGCCGTGCCATGTCAATGA---AGATCTCTTGGCTTTGGCTGTGGAAGATTTTACGCTTT 1578
QY 1563 TGCCAATCTATTATTCGTGGAAGAAATTAAGAGTCTTTGAAAGGTGGTGGTGTAGAGAAATAG 1622
Db 1579 TGTCAGTCAACTTACCAAAATGAATCCAGCACCTTTGAAAGTGGGAGAAAGAAATTAAG 1638
QY 1623 TTGACACGAGCTCAAGTTTTCGTAGGCAAAAGACCGCTACTGTTATTTCTCTGTGCTGCA 1682
Db 1639 CTGACCAAGCTCGAATTTTACGCGGAAGAAATCTAATAAACAGCTATCTCTCTGCTGCTGCC 1698
QY 1683 ACATTTTCTCTCCGAATTTATCAGATGCGGTATTTTTCATGGGCCCAAAATTTGGCATATTA 1742
Db 1699 ACCATAGCCCTTATGAATTTCTGATGCTGCGATTTGCGTGTGCAAAATCTATTGCGCTC 1758
QY 1743 ACTACAGTAGTTGATGACTTTTTTTCATATCGGTGGTGAACAATCGATGAATTCACCAACTG 1802
Db 1759 ACATTTGTTGCCGATGACTTTTTTCGATGTCGGAAGTTCCAAAGAGAAACAAGAAATCTC 1818
QY 1803 ATTCATGTTTGAATAATGAATGTAGATGTCGAACAGGATTTGTTTCAGAGCATGTT 1862
Db 1819 ATATCTTTAGTCGAGAAAGTGGGATCAGTATCACAAAGTTGAGTTCTACTCTGAGAAATGTA 1878
QY 1863 CGGATTTTATTTTATGATTAAGATTAAGATGCAATCTGTTTCAAACTTGGTTGGAATTAATGAAT 1922
Db 1879 AAGCAGTATTTTTCGCTCTATATTTTACGGTTAAACAGCTTTGAGCAATGGCTTCCGCA 1938
QY 1923 TGGCAACGCGCGAGTAACTAGCCATGTTTATTTCAAACTTGGTTGGAATTAATGAATAGT 1982
Db 1939 GTACAGAACCGCAGCGTTACAAATAACAATCTGAAATCGTGGTTGGATTTATTTGAGTCT 1998
QY 1983 ATGTTGAGAGAGCTATATGGAACAGAGATGCTTATGTCACCAACATTTAAATGAATATATG 2042
Db 1999 TTACGACAGATGAGAAATGGCAACGAGCAAAATATGTGCCAACATGGAGGAATACATG 2058
QY 2043 GAAAACCTTTACGTTGCTCATTTGCAATTTAGCCCGATTTCTCAAGCCGCTATTCTTTGTG 2102
Db 2059 AAAAAATCAATTTGACATTTCCGATTTGGGACCAACTATATCTATGACACTGTATTTTCATG 2118
QY 2103 GGGCCCAAAATATCAGAGGAGATTTGTTGAAGCTCTGAATATCATATCTATTTTAAAGCTA 2162
Db 2119 GGAACAAATCTCTGGGAGGACATCGTGAAATAATGCAGAGTATGATGATGTTGTTAGACTA 2178
QY 2163 ATGAGCAGCAGAGGTGCACTTTCTAAACGATATCCATGCTTCAAGAGGGAAATTTAAGGAA 2222
Db 2179 ATGAACACATGTTGCTCGTCTCCAGAAATGATTTCAAAGCTTTTGAAGAGGAATGCAAGGAT 2238
QY 2223 GCAAAATTAACCGCGGTAGCATTTGAGTAAACGAGAAAGTGGAAAGTGGAGAA 2282
Db 2239 GGCAACTGACAGTGTGTCACTGTTGTTCTTGAACAGAAAGATGTCATGTCTAGTAGAA 2298
QY 2283 GAGGTTGTGGAGAGATGATGATGATTAATAAACAAGAGGAAGAATTAATGAATTA 2342
Db 2299 GAGGCTAAAGAGCGGATAAACGAGTCTATATCATCATGTAGAGAGAGATTTGCTACGGTTG 2358
QY 2343 ATTTTGAAGAAATGTTAGCATTTCTTAGAGCTTTGTAAGATGCAATTTTGGAAACATG 2402
Db 2359 GTTGTGTAGAGAGACGGTGTCA---TTCTTAATCATGCAAGGAGATGTTCTGGAATCTT 2415
QY 2403 TGTCACTGTGTGAATTTTTCACGCAACGATGACGGGTTTAC 2446
Db 2416 TACAAGCAAGCAATGTGTTTACTCTCAGGCGGATGGATTTTC 2459

RESULT 13
AK119327

LOCUS

AK119327 2753 bp mRNA linear PLN 29-OCT-2003

Oriza sativa (japonica cultivar-group) cDNA clone:001-131-B02, full insert sequence.

ACCESSION AKI19327
 VERSION AKI19327.1 GI:37988950
 KEYWORDS FLI CNA; oligo capping.
 SOURCE Oriza sativa (japonica cultivar-group)
 ORIGIN Oriza sativa (japonica cultivar-group)
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

1
 The Rice Full-length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team:
 Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuka, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Ootomo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nishikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN:
 Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.
 Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

TITLE japonica rice
 JOURNAL Science 301 (5631), 376-379 (2003)
 MEDLINE 22752273
 PUBMED 12869764
 REFERENCE 2
 AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Kato, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Negata, T., Nakahama, Y., Nakamura, M., Namiki, T., Narikawa, R., Nishikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuka, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Ootomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shingawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S., and Yoshimura, A.

Collection, mapping, and annotation of 28K full-length cDNA clones from japonica rice
 Unpublished
 3 (bases 1 to 2753)
 Kikuchi, S.
 Direct Submission
 Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp, Tel: 81-29-838-7007, Fax: 81-29-838-7007)
 This clone is one of the 32K full-length cDNA clones from japonica rice.
 URL: <http://cdna01.dna.affrc.go.jp/cdna/>
 NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuka, K., Shishiki, T.,

Yamamoto, M. and Nakahama, Y.
 FAS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Nishikura, J., Nishi, K., Ooka, M., Ryu, R., Sugano, S., Sugiyama, A., Matsubara, K. and Murakami, K., Xie, Q., Yokomizo, S., Yoshimura, A., Tsunoda, Y., Ueda, M., Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Kato, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sato, K., Shibata, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.
 Location/Qualifiers
 1..2753
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="001-131-B02"

FEATURES
 source
 ORIGIN
 Query Match 16.7%; Score 467.6; DB 8; Length 2753;
 Best Local Similarity 53.7%; Pred. No. 1.9e-81;
 Matches 1184; Conservative 0; Mismatches 959; Indels 61; Gaps 8;
 QY 285 ACTATCTGATCATTCATACACAAAGAACCGATCCAAAACAGCTTTAAAATGTAGAA 344
 DB 333 ACAAGTCAGATAAACAATGGGATGGGAGCTAGATACTTAGGCATCTGGAGAACCGGAG 392
 QY 345 ATTCTCTGTTTCTTCATATGACACAGATGGGAGCATGGTCCCTTCTCCAACTCA --- 401
 DB 393 TTCTTACCATCTTCATATGACATAGCAGUGGGTGGTATGGTCCATTCGCGGCATGAT 452
 QY 402 ---CCCAATCGCTTGTGTTCCCTCGATGTTCTCAATGGTGTATTAATTAATCAGCTTAAT 458
 DB 453 CATCTTCAGCTCCATGCTTCCTCGAATGTGTGAATGATGATCTACAAAACCAACACAGT 512
 QY 459 GATGGTTCTAGGGGCTTGTGTTTAACTACACTCATATCATATCACCCTGTTGCTTAAAGAT 518
 DB 513 AATGGGTCTGGGGGTGTCA-----ATGAATTTGACTCATCAGCCAGCAAGGAT 560
 QY 519 TCTCTATCTTCAACATAGCATGATTTGTTCATTTAAAAGATGGAATGTGGGGAAGAT 578
 DB 561 ATTCTCTTCACTTCGTCATGTTATTTGACATTCGACATTCGACATTCGACATTCGCGTCCGAG 620
 QY 579 CAAATAAATAAAGGCTAAAGTTTATTTAGTCAATCTTCCTTCAGCTACTGAAAAGAT 638
 DB 621 CAAATAAGGAGAGGAGTATACATTTTATCGCAAGATTTCTCCATTTGATGATGCCAG 680
 QY 639 CAACCATCTCCCATTTGGTTTTCATCATATTTTCTCGTTTGTCTGATGATGCAAAAAC 698
 DB 681 ATTGCTGACCTATAGGCTTCACCTTCACATTCCTTCGTTTAACTTGCATTAAG 740
 QY 699 TTGGACATATAACCTCCTTTTCAAAACAAACAGATTTTGTAGTGTACATAAGAGGAA 758
 DB 741 ATGGGTTTGAATTTCTTCGCTAGGAAATTTAGTATTGATCAGATTTCTTCACTCCGCTGAT 800
 QY 759 TTGAGCAAAAAGAG-----TCCCATTCAAATGAGATGATGATGATTTGGCGTAT 809
 DB 801 ATGGGTTTGAAGAGCTGGCTGGTGGAGCAATTTTGGGAAAGAGGAGATATTTCCGCTAT 860
 QY 810 ATCTCTGAAGACT---CGGTAATTTATATGATTGGAATATGTTGAAGAATATATCAGATG 866

QY 633 AAAAGTCAACCATCTCCCATTTGGTTTGGACATCATATTTCTGCTGTTTCTGTTAGTATCG 692
Db 192 GAGCAGTTCACTTCTCTATAGGTTTCAACTTCACTTCTCTGTTTCTTAGCCTCGGC 251
QY 693 AAAAATCTGGACATAAAACCTCTTTCMAAAACAAAACAGATTTTGTGTTGATGCTACATAAG 752
Db 252 ATTGATATGGGTTAGATTTTCTGTGAAGACAAATTTGATGCTGTGGCATTTCTTCAACGC 311
QY 753 AGGGAATTTGGAGCAAAAAGAGTGC-----ATTCAATGAGATGAGTGGATCTTG 803
Db 312 CGGAGATGGAATTTGAAAAGGCTGCTGTGGATAGTTCTTTTGGAGAAAAGCATATAG 371
QY 804 GCGTATATCTCTGAGAGGCTCGTAAATTTATATGATTTGCAATATGCTGCAAGAAAATATCAG 863
Db 372 GCTTTTATCCCAAGAGGATTCGAAATATGCTGGATGGAATCAAGTTATGAAGTTTCAG 431
QY 864 ATGAAAAATGTTCTGTTTTCACCTCACCATCAGCAACAGCTGCTGCTTTTCAATTAATCAT 923
Db 432 AGGAAGATGGATCATTTGTTAGCACTCTTCCACAACTGCTGTTGCAATTAATCCACAAA 491
QY 924 CAAAATCTCGTTGTTCTTAATTTATTTAAATTCATTTTGGACAAGTTTGGTAATCAGTC 983
Db 492 TACAACGACCAAGCCCTTCAATACCTAAATTTGCTTGTCAATGAATTTGGCAGTGAGTA 551
QY 984 CCAACAGTTTATCCTCATGATTTATTTATCCGACTTTCATGTTTGACACAAATTTGAAAGA 1043
Db 552 CCAGCAATGATATCCTTCAAGGATCAATTTGAGCTTTCAATGTTGGAACGCTTGAATAA 611
QY 1044 TTAGGAATTTCAACCATTTCAAGTGGAAATTTAAATTTTATAGTGAATCAATACAGA 1103
Db 612 ATGGGAATTTCTCAGCGCTTTGTCAGTGAATAGAAAGCATCTGCAATGCGCATACAT 671
QY 1104 TGTGGGTGGAACGAGATGAGCAATATTCATGATGTTTGAACATGTTTGTAGCTTTT 1163
Db 672 TGCTGTTACAGAAATGATGAGGAATCATGATGGACATAGCAACATTTGCAATGGCATTT 731
QY 1164 CGGTTATTAAGGATCAATCGGTATGAAGTTTCCAGATCCATTTGGCTGGAATTAATAAT 1223
Db 732 CGCCTTTTGAGGATGAATGTTACGATGTTTCTCAGATGAGCTGTCTCAGCTTGTGGA 791
QY 1224 G-----AATTAGCTTTCAAGAGCAATATGAGCTCTTGAA 1259
Db 792 GCTTCCACTTTCCATGATTCACATACAGGATATTTAAATGATACAAATCCCTACTTGAA 851
QY 1260 ACATATCATGCTGACAC---TATATATACAGAGATTTATCTTCTGGAACAAATC 1316
Db 852 TTGTACAAGACCTCAAAAGTCACTTATCAGAAACGATCTGATCTTATGATCGCATAGGT 911
QY 1317 TTGAAGTCTAGCTGATTTCTCAAGAGATATATCCAC----- 1354
Db 912 TCCTGTTCTGGCAACTTATTTGAAGATAGATGCTGTAGTAAGGTGCAAAAGACTCGA 971
QY 1355 -----TGATTTCAACAGGCTTTCTAAATTAATTCAC----- 1386
Db 972 TTTTGTGAGAGATGCTGCAACAAATAATTTAAATTTCTCATTTTCACTTTGGAGGTTT 1031
QY 1387 -----RAGAGTGGAAATGCTCTTAAGTTCCCTATCAATACCGGT 1427
Db 1032 TATGCTCTGTTTGTGTTTTCAGATCGATGATGCTGTTAAATTTTCCCTTGTATTCACA 1091
QY 1428 TTAGAAGCATAAACACTAGACGAAATATACAGCTTTTACAAATGTAGACATAACAGAAAT 1487
Db 1092 CTGGAGGCTTAGACACACAGAGAAACATCGAACATTTTGAATGCTTTGGGTTTCTCTGATG 1151
QY 1488 CTGAAACATCATATCATCATCAAAATTTAGTAACTGATTAACCTGAAGTTTGGCTGTT 1547
Db 1152 CT---AACAACAAATCTCATCTTTTCTGATCAATCAAGAAATTTCTAGCTTTGGCAGTC 1208
QY 1548 GAAGATTTCTACACCTCCCATCTTATTTATCGTGAAGAAATTTAAAGCTTTGAAAGGTGG 1607
Db 1209 GAAGATTTCTAGTTTCTCTCAACGTTTACCGGATGAATCTCGGATCTTGTAGTTGG 1268
QY 1608 GTGGTAGAATAAGTTGGACCACTCAAGTTTGTAGGCAAAAGACCGCTACTGTAT 1667

Db 1369 GTGAAGGAGAAACAGCTGGACCACTACAAATTTCTCGGCAGAAACTGACATATTTCTAT 1328
QY 1668 TTCTCTGTGTGCTGCAACACTTTTCTCCCGAATTTATCAGATGCGGTATTTTCATGGGCC 1727
Db 1329 CTGTCTGTCTGCTGCTACCGTATTTTCTCTGAATTTGCTGAGCTGCGCATTTTCATGGGCC 1388
QY 1728 AAAAATGCAATTAACACTACAGTAGTTGATCACTTTTGTATATGCGTGGTACAAATCAT 1787
Db 1389 AAAAATGCTGCTTCCACAACTGTGTTGATGACTTCTTCTGATTTGTTGGTATCAAAAGAA 1448
QY 1788 GAATGACCAACCTGATTTCAATGTTTGAATAAATGGAATGATGTTGCGACAAAGATTTGT 1847
Db 1449 GAAATAGAAAACCTGATAGCACTAGTTGAGAAATGGCATGGGCAACCATGCTAGTTGAGTTC 1508
QY 1848 TGTTCAGAGCATGTTTCGATTTTATTTTAGCATTTAAAGATGCAATCTGTTTGGATTGGA 1907
Db 1509 TATTCGGAACAGGTGAAATATGATTTTCTGCTATTTATACAACTGAAACCATCTTGA 1568
QY 1908 GATGAAGCTTTTAAATGCGAAGCGCGCATGTAATAGCTAGTCTGTTTATTTCAAACTTTGGTTG 1967
Db 1569 GCAATGGCTTCTGACAGCACAAAGCGCTGATCTTACAAACCACTAGTAGAATAATATGGCTG 1628
QY 1968 GAACTAATGATAGTATGTTGAGCAAGCTATATGCAACAGAGATGCTTATGTCACAACA 2027
Db 1629 GATTTGTTAAGATCTATGATGCTGAGGCGAAGATGGCAGAGATGCCAATATGTACAAACA 1688
QY 2028 TTAATGATATATGTAATAACGCTTTACGTTGTCATTTGCAATTAGGCGCCGATTTGCAAGCGC 2087
Db 1689 GTTGAAGATATACATGACAAATGCTGTTGCTCATTTGCACTGGGCCCAATTTGCTGCCCA 1748
QY 2088 GCTATTTTACTTTTGGGGCCCCAAATTTATCAGAGGAGATTTGTTGAAGCTCTGAATATCAT 2147
Db 1749 GCATTTGATTTTGTAGGCAAGAGCTATTAGAGCATGCTGTCAAAGATGAAGAGTAGCAT 1808
QY 2148 AATCTATTTAAGCTAATGAGCAGCAGGTCGACTTCTTAAACGATATCCATAGCTTTCAAG 2207
Db 1809 AAATATTTTAGCTAGTGAGCACTTTCGGGAGGCTTCTCAATGACTACCAAAAGTTTAGAG 1868
QY 2208 AGGGAATTTAAGGAAGGCAATTTAAACGGGTAGCATTTGCATTTTGTAGTAAACGAGAAAGT 2267
Db 1869 AGGGAAGCAACAGGGGAGCTGATAGTTGTTCTTACTTGTCTCCACAGTGTGGT 1928
QY 2268 GGGAAAGTGGAAAGAGAGGTTGTGGAGAGATGATGATGATGATTTAAACAGAGAGAA 2327
Db 1929 TCTATGCTCATAGAACCCGCTAAAGGCAATGCAAGAGTCCATAGACGCTGTCTAGAGA 1988
QY 2328 GAATTAATCAAAATTAATTTTGAAGAAATGTTGATGATTTTCTAGAGCTTCTAAAGAT 2387
Db 1989 GACTTGTAGATTTGTTCTT---CAGGAAGAAAGTGTGTTCTTCTAGGCCATGCAAGGAG 2045
QY 2388 GCATTTTGGAAATGTCACGTTTGAATTTTAAATTTTAAAGCAACGATGACGGTTT 2445
Db 2046 CTCCTCTGGAAGATGTTAGATACTTCCCTGTTTACTCTCAGAATGATGATTTA 2103

RESULT 15

AK068310 3061 bp mRNA linear PLN 24-JUL-2003
LOCUS Oryza sativa (japonica cultivar-group) cDNA clone: J013149A01, full
DEFINITION insert sequence.

ACCESSION

AK068310
VERSION AK068310.1 GI:32978328

KEYWORDS

FLU_CDNA; CAP trapper.

SOURCE

Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)

ORGANISM

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzaceae; Oryza.

REFERENCE

1
The Rice Full-Length cDNA Consortium, National Institute of
Agrobiological Sciences Rice Full-Length cDNA Project Team,
Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,

Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Nami, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Ohtsuki, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN: Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saitoh, K., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.

Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

Science 301 (5631), 376-379 (2003)

2752273

12869764

2 (bases 1 to 3061)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, N., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Nami, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Ohtsuki, K., Ohtsuki, K., Saitoh, K., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Shishiki, T., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, T., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S., and Yoshimura, A.

Direct Submission

Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp, Tel: 81-29-838-7007, Fax: 81-29-838-7007)

This clone is one of the 28K full-length cDNA clones from japonica rice.

URL: <http://cdna01.dna.affrc.go.jp/cdna/>

NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Sato, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Nami, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., and Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Ohtsuki, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K., and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, Y., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,

Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A., and Hayashizaki, Y.

Location/Qualifiers

1..3061

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="mRNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/clone="J013149A01"

FEATURES

source

ORIGIN

Query Match

Best Local Similarity 16.1%; Score 448.2; DB 8; Length 3061;

Mismatches 1156; Conservative 0; Mismatches 963; Indels 60; Gaps 7;

308 CAAAGACGGATCCAAACACAGTTTAAATAATGTAGAAATTTCTGTTCTTCATATGACAC 367

414 CGAGCGAGTATAAGGAAGCAGCTCCAGGAGTGAATATCAACATCCCATATGACAC 473

368 AGCATGGGTAGCCATGGTCCCTTCTCCAAATCTCACCCTAAATCGGCTTGTTCCTGAGTG 427

474 TGCATGGGTAGCTATGGTGGCCACTCCGGGGTTCGTCTCACAATCCAAGCTTCCCTCAATG 533

428 TCTCAATTTGGTTAATTAATATACAGTTAATGATGGTTCATGGGGTCTTGTGTAATCAAC 487

534 CGTCGACTGGATCTAGAGAAATCAATGGGACGATGGATCATGGAGTATTTGACGGGTCCAT 593

488 TCATPAATCATATCAATCAATCCCGTTGCTTAAAGATTTCTCTATCTTCAACATTAGCATGTATTGT 547

594 ATCCACAGCCAA-----CAAGATGCTCTATCATCTAGCTAGCATGTGTTCT 641

548 TGCATTTAAAGATGGAATGTTGGGAGAGATCAATAAATAAAGGTCTAAGTTTATTGA 607

642 GGCCTCAACAAATGGAATGTTGGTAGGAGCACATCAGGAGAGGACTAAGTTTCATTGG 701

608 GTCAAAATCTTGTCTACGCTACTGAAAGAGTCAACCATCTCCCATTTGGTTTGCATCAT 667

702 AAGAAATTTCTCGATTCGGATGGATGATCAAGCAAGTTGCTCCTATAGTTTTCGGCATCAC 761

668 ATTTCTCGTTTGGTTGAGTATCGAAAAAATTTGGACATAAACTCTCTTTCAAAACAAAC 727

762 TTTCCAGCTATGCTAACCCCTTGCCATGGTTCGGCTTCGGAGTCCCGCTCAGACAAA 821

728 AGATTTTGTGATGCTACATAGAGGGAATTTGGACAAAAAAGATG-----CCA 778

822 CGATTTGATGATGCTTTAATCACTTCGGAGAGTAAAAATACAAAGAGAGCGCGGAACCA 881

779 TTCAAAATGAGATGGATGCTTGGCTATATCTCTGGAAGGACTCGGTAATTTATATGA 838

882 TTCTCGTGGAGAAAGCCCTATATGGCTATCTGGCAGAAAGGATTTGGGAACCTTCTGGA 941

839 TTGGAATATGGTCAAGAAATATCAGATGAAAAATGTTTCTGTTTCAACTCACCATCAGC 898

942 GTGGATGAAATCATGATGTTCCAGAGGAAGAACGGGTCTATTGTTCACTGTCTTCTTC 1001

899 AACAGCTGCTGCTTTTCAATTAATCATCAAAATCTCGTTTGTCTTAATTTAATTAATCACT 958

1002 AACTGCTGGCGCATTAGCCAAATTTACCACGACGATAAAGCTCTCCAATCTTGCATCTCT 1061

959 TTGGCAAGTTTGGTAAATGTCAGTCCCAACAGTTTATCTCTCATGATTATTATTCGACT 1018

1062 AGTCAATAATTTGACGGTGTAGTGCACACATGTAATCCATAAATATTTTGTGAGCT 1121

1019 TTCTATGTTGACACAAATTTGAAAGATTAGGAATTTTCAACACATTTTCCAGAGTGAATAA 1078

1122 TTCAATGTTGGATGCGCTTTGAAAAATGGAATATCTCAGTATTTTCCAGTGAATAA 1181

1079 AAATGTTTGTAGTGAACATACAGATGTTGGTGGAAACGAGATGACCAATATTTATGGA 1138

1182 GAGCATCTTGCATGACATACAGATTCTCTGGTTAGGAAAGATGAGGAAATCATGTTAGA 1241

1139 TGTGTTAATCATGCTTTTACGCTTTCGGTTATTAGGATCAATGGTATGAAGTTTTCCTCC 1198

Mon Oct 18 07:21:08 2004

Db	1242	CGTAAACAACGTGCGCAATTTGCGCTTTTAAAGATGAACGGTTACGATGTTTCCCTC	1301
QY	1199	AGATCCATTGGCTGAAATTTACTAAATGAATTAG-----CTTT	1234
Db	1302	AGATGAGCTATCTCATGTTTGTCTGGAGCTTCAGGCTTCGGTATTCGCTTCAAGGGTATCT	1361
QY	1235	GAAGAAGCAATATGCAAGCTCTTTGAACAATATCATCATCGCTC---ACATATATTATACCAAGA	1291
Db	1362	AAACGATAGAAATCTGTACTGGAAGTATACAAGAAGCTCGAAACACACGATATACAGAAA	1421
QY	1292	GGATTTATCTTTGGAAGAAACAAATCTTGAAGTCAGCTGATTTCTTCAAGAGAGATAATATC	1351
Db	1422	TGATTTTGATCTTTAGATAGCATAGGCTCTTGGTCAGGCAGCTATTGGAAGGAAATGCTGTG	1481
QY	1352	CACGTATTCAAAACAGGCTTTCTAAATTAATTCACAAGAGGTGGAAAATGCTCTTAAAGTT	1411
Db	1482	CTCTAATGGGAAGGAATCTCC-----AGCCGTGAGAGATCGAGTTTGCTCTGAAGTA	1535
QY	1412	CCCTATCAATACCGGTTTGAACGCAATAAACACTAGACGAATAATACAGCTTTTACAATGT	1471
Db	1536	TCCCTTCTTACAGCACATTTGAACGCTAGTCCACAGGAAGAACATTTGCTCTTTCCAGCGC	1595
QY	1472	AGACATACAGAAATTTCTGAAAACACTACATCACTCATCAATATTAAGTAAACACTGATTA	1531
Db	1596	TAAAGGCTCCAGATGTCTGAAAACAGAGTGCAATGCGCTTTCATGACAGTCA---AGATTT	1652
QY	1532	CCTAAGGTTGGCTGTTCAAGATTTCTACACCTGCCAATCTATTATTCGTGAAGAAATTAAA	1591
Db	1653	TCCTAGCTTTGGCTGTGATGATTTCTGCACTCTCAATCTAATTACCGAATGAATCTGAA	1712
QY	1592	AGGCTTTGAAGGTGGGTGAGAGATAATAGTTGGAACCAAGCTCAAGTTTGTAGGCAAAA	1651
Db	1713	TTATCTCGAAGTTGGGTGAAAGACACAGACTCGACCAAGCTACATTTGCGACGGCAGAA	1772
QY	1652	GACCGCTTATCTGTTTATTTCTGTTGCTGCAACACATTTGCTCTCCGGAATTAACGATGC	1711
Db	1773	GATAACATATTGCTATCTCTGCTGCTGCCACCAATTCGCTGCTGGAATGGGCTACGCG	1832
QY	1712	CGGTATTTTCATGGGCCAAAATGSCATATTAACCTACAGTAGTTTGATGACTTTTGTGATAT	1771
Db	1833	TCGCACCTCTGTGGGCAAGAACCGCTTGCTGTTGAACGGCTGTTATCGACGATCTCTTCACT	1892
QY	1772	CGGTGTACAAATCGATGAATTTGACCAACCTGATTTCAATGTGTTGAAAAATGGAATGTAGA	1831
Db	1893	TGTTGGATTAGAACACAGAAACAAAGAAACCTCTTAGCATTAATGAGAAGTGGGAAGAGCC	1952
QY	1832	TGTGCAAGAAGATTTGTTGTTTCAGAGCATGTTTCGGAATTTTATTTTATAGCATTAAGAATGC	1891
Db	1953	TGGTGAAGATGAGTACTACTCTGAAGATGTAAAGATCGTGTCCAGGCTCTTTATTAATAC	2012
QY	1892	AATCTGTTTGGATTCGAGATGAAGCTTTTAAATGGCAAGCGCGATGTAAGTAACTAGCCATGT	1951
Db	2013	GGTGAACAGATTTGGCCAAAGGCTTTCGGGTTCACAGGCCATGATTTACCAAGTACCT	2072
QY	1952	TATTCAAAACCTTGGTTGGAACATAATGAATAGTATGTTGAGAGAAGCTATATPGGACACAGAG	2011
Db	2073	AGTAGACGTATGGCTACTCATGTGGTGAGGTGTATGAAGGTTTGAAGCAATGGCAGAGGAG	2132
QY	2012	TGCTTATGTGCCAACATTAATGAATATATGGAAGAACGCTTACGTGCTATTTTCATTAGG	2071
Db	2133	CCAGCATCTGCCAAACGTTTCGAGAGATACATGAAAGTGGCATGTGTGCTCTTGGACAGGG	2192
QY	2072	CCCGATTTGCAAGCCGCTATTTTACTTTGTTGGGCCCAAAATATATCAGAGAGATTTGTGTA	2131
Db	2193	CTGTACCGGTGATGTCCGCAATTTGTTCTGATCGGAGAAAAGCTCCCGAGGGTATTTGTCGA	2252
QY	2132	AAGCTCTCAATATCAATATCTATTATTAAGCTAATGAGCAACGACGGGTGCACTTTCTTAAACGA	2191
Db	2253	ACTTGAAGAAATACGATGAGTTGTTTCAGGCTGATGGGCACTTGCGGCCGTCTCTCGTAATGA	2312
QY	2192	TATCCATAGCTTCAGAGGGAAATTTTAAGGAAGGCAAAATTAACCGCGGTAGCATTTGCAATTT	2251
Db	2313	CAATTCAGGCAATTTGAGAGGAGGAGTTCGGAACGGCAAAATGACGAACGGCGCTCTCGCTGCT	2372

Qy	2352	GAGTAA	CGGAGAAAGTGGGAAAGT	---GAGAAACAGAGGTTGTGGAGGAGATCATGATGAT	2308
Db	2373	CGTTAC	CGCTAGCGCGCGCTCCATGTC	CGTACGAGCGCCAAACCGAAGTGTGNAAGCG	2432
Qy	2309	GATTAAAA	CAAGAGGAAGAAATTAATGAAATTAATTTTTGAGAGAAATGTTAGCATTTG	2368	
Db	2433	CATCGAGCCCTAA	CGGAGGAAGTCTCTGAGCTTGGTTCGTACGGAACAGGAAGGCCCTAT	2492	
Qy	2369	TCCTAGAGCTTGTAAAGATGCATTTTTGGAACATGTGTACGTTGTGAAATTTTTTTTACGC	2428		
Db	2493	CCCGAGGCGGTGCACAGCAGCTCTCTGGAAGATGTGCAAGATCTTCACCTGTTCTACTA	2552		
Qy	2429	AAACGATGACGGTTTTACT	2447		
Db	2553	CCGAGCCGACGGATTCACT	2571		

Search completed: October 16, 2004, 23:04:19
Job time : 11760.6 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 03:30:51 ; Search time 124.839 Seconds
(without alignments)
1543.991 Million cell updates/sec

Title: US-10-041-018-22

Perfect score: 1720

Sequence: 1 MEAKIDELINNDPVWSSQNE.....DTATNHLDELVIIDHLSL 335

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1720	100.0	335	2	Q12051	Q12051 saccharomyc
2	838	48.7	332	2	Q6FRW9	Q6FRW9 candida gla
3	753	43.8	320	2	Q758K0	Q758K0 ashbya goss
4	753	43.8	320	2	RA552447	RA552447 ashbya go
5	634	36.9	396	2	Q87411	Q87411 emericella
6	629	36.6	327	2	Q6C8T9	Q6C8T9 yarrowia li
7	628.5	36.5	418	1	GGPP GIBFU	GGPP GIBFU g geranylge
8	625.5	36.4	375	2	Q9C452	Q9C452 penicillium
9	623	36.2	428	1	GGPP NEUCR	GGPP NEUCR n geranylge
10	614	35.7	387	2	Q6MVH1	Q6MVH1 aspergillus
11	614	35.7	387	2	CAF32032	CAF32032 aspergill
12	612.5	35.6	343	2	Q6F8E2	Q6F8E2 phoma betae
13	598.5	34.8	343	2	Q6F5E6	Q6F5E6 phoma betae
14	591.5	34.4	303	2	Q9P885	Q9P885 mucor circi
15	583	33.9	300	1	GGPP HUMAN	GGPP HUMAN h geranylge
16	581	33.8	300	2	Q6NWT9	Q6NWT9 homo sapien
17	581	33.8	300	2	RAH67768	RAH67768 homo sapi
18	578	33.6	294	1	GGPP BOVIN	GGPP BOVIN m geranylge
19	577	33.5	300	1	GGPP MOUSE	GGPP MOUSE m geranylge
20	577	33.5	300	2	AAH69913	AAH69913 mus muscu
21	576	33.5	315	2	Q6CMK36	Q6CMK36 kluyveromyc
22	570	33.1	300	2	Q6F596	Q6F596 rattus norv
23	558.5	32.5	327	2	Q7ZHY0	Q7ZHY0 brachydanio
24	554	32.2	392	2	Q9P965	Q9P965 nigrospora
25	553	32.2	371	2	Q9C446	Q9C446 penicillium
26	545.5	31.7	338	2	Q61539	Q61539 drosophila
27	545.5	31.7	338	2	Q9V854	Q9V854 drosophila
28	541.5	31.5	346	2	Q6BN82	Q6BN82 debaryomyce
29	527	30.6	335	2	Q86F69	Q86F69 schistosoma
30	524	30.5	236	2	Q9CZ26	Q9CZ26 mus muscula
31	513	29.8	335	2	Q61538	Q61538 drosophila

32	500	29.1	357	2	Q7QI19	Q7qii9 anopheles g
33	454.5	26.4	342	2	Q74289	Q74289 gibberella
34	337	19.6	347	2	Q51294	Q51294 borrelia bu
35	310.5	18.1	327	1	IDS_A_METJA	IDS_A_METJA methanococ
36	282	16.4	323	2	Q8D276	Q8d276 wigleswort
37	282	16.4	324	2	Q7VQN3	Q7vqn3 candidatus
38	274.5	16.0	364	2	Q05708	Q05708 treponema p
39	268	15.6	350	2	Q73MH1	Q73mh1 treponema d
40	268	15.6	350	2	AA512054	AA512054 treponema
41	262	15.2	317	2	Q6M172	Q6m172 methanococ
42	262	15.2	317	2	CAF29601	CAF29601 methanoco
43	249	14.5	742	2	Q9NEB3	Q9neb3 leishmania
44	246.5	14.3	322	2	Q88Q11	Q88q11 pseudomonas
45	244	14.2	324	1	IDS_A_METTM	IDS_A_METTM m bifunctio

ALIGNMENTS

RESULT 1

Q12051	PRELIMINARY;	PRT;	335 AA.
AC Q12051			
DT 01-NOV-1996	(Tremblrel. 01, Created)		
DT 01-NOV-1996	(Tremblrel. 01, Last sequence update)		
DT 05-JUL-2004	(Tremblrel. 27, Last annotation update)		
DE Btslp: geranylgeranyl diphosphate synthase (Btslp).			
GN Name=BTS1;			
OS Saccharomyces cerevisiae (Baker's yeast).			
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX NCBI_TaxID=4932;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TRANSPOSON=Ty4;			
RX MEDLINE=97313271; PubMed=9169875;			
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W.,			
RA Araujo R., Aparicio A., Barrell B., Badcock K., Benes V., Botstein D.,			
RA Bowman S., Bruckner M., Carpenter J., Cherry J.M., Chung E.,			
RA Churcher C., Coster F., Davis K., Davis R.W., Dietrich F.S.,			
RA Delius H., DiPaolo T., Dubois E., Dusterhoft A., Duncan M., Floeth M.,			
RA Fortin N., Friesen J.D., Fritz C., Goffeau A., Hall J., Hebling U.,			
RA Heumann K., Hilbert H., Hillier L., Hunnicke-Smith S., Hyman R.,			
RA Johnston M., Kalman S., Klein K., Komp C., Kurdi O., Lashkari D.,			
RA Lew H., Lin A., Lin D., Louis E.J., Marathe R., Messenguy F.,			
RA Mewes H.W., Mirtipati S., Moestl D., Muller-Auer S., Namath A.,			
RA Nentwich U., Oefner P., Pearson D., Petel P.X., Pohl T.M.,			
RA Purnelle D., Schafer M., Scharfe M., Scherens B., Schramm S.,			
RA Schroeder M., Sdicu A.M., Tetelin H., Urrestazu L.A., Ushinsky S.,			
RA Vierendeels F., Visers S., Voss H., Walsh S.V., Wambutt R., Wang Y.,			
RA Wedler E., Wedler H., Winnett E., Zhong W.W., Zollner A., Vo D.H.,			
RA Hani J.;			
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";			
RL Nature 387:0-0(0).			
RL [2]			
RN SEQUENCE FROM N.A.			
RC TRANSPOSON=Ty4;			
RA Winnett E., Ahmed A., Bussey H., Fortin N., Friesen J.D., Hall J.,			
RA Storms R.K., Vo D.H., Wang Y.;			
RN Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.			
RL [3]			
RP SEQUENCE FROM N.A.			
RC TRANSPOSON=Ty4;			
RA Jia Y., Cherry J.M.;			
RN Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.			
RL [4]			
RP SEQUENCE FROM N.A.			
RC MEDLINE=95394944; PubMed=7665600;			
RA Jiang Y., Proteau P., Poulter D., Ferro-Novick S.;			
RT "BTS1 encodes a geranylgeranyl diphosphate synthase in Saccharomyces			
RT cerevisiae.";			
RL J. Biol. Chem. 270:21793-21799 (1995).			
DR EMBL; U9205; AAB68296.1; -.			


```
GN ORFNames-YALI0D17050g;
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RG GENOLEVURES;
RA DuJon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talia E.,
RA Goffard N., Frangin L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boissame A., Boyer J., Cattolico L., Confaniolieri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenlou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR EMBL; CR382130; CAG81114.1; -.
SQ SEQUENCE 327 AA; 37316 MW; A982D91858A2F302 CRC64;

Query Match 36.6%; Score 629; DB 2; Length 327;
Best Local Similarity 40.8%; Pred. No. 4.8e-36;
Matches 131; Conservative 62; Mismatches 104; Indels 24; Gaps 6;

QY 14 WVSSQNESLISKPNHLLKPGKFNRLNLIQVNRVNNLPKQDLATVSOIVELLHNSLL 73
DB 11 IWGKAADTALLGPNVLANRGNINREHLIAFGAVIKVDSLETISHITKLHNSLL 70

QY 74 IDIEDNAPLRGQTTSHLIFGVPSTINTANYMFRAMQLVSQLTTKEPLYHNLITFNE 133
DB 71 VDDVEDNSMLRGLPAHCLFGVPQTINSANYMFVALQELKLS- - - -YDAVSIFTE 125

QY 134 ELINLHRGQGLDIYWRDPEIPIPTQEMLYNMVNKTGGLFRILRLM-EALSFSHHGH 193
DB 126 EMINLHRGQGLDIYWRDPEIPIPTQEMLYNMVNKTGGLFRILRLM-EALSFSHHGH 182

QY 194 LVPPF- - - -INLIGLIYQIRDDYLNKDFQMSSEKGFPAEDITEGKLSFPPIVHALNFTKTK 248
DB 183 KINFDLTHLTDTLGVIIYQILDDYLNKDFQMSSEKGFPAEDITEGKLSFPPIVHALNFTKTK 237

QY 249 GQTEQHNEILRIILLRTSDKDIKLIQILEFPTNSLAYTKNFINOINMKNENKYL 308
DB 238 -TNPDNHEILNLIKQRTSDASLKKYADVYWRTEKTSFDYCLKRIQAM- - - -SLKASSYI 231

QY 309 PDLASHSDPATNLDHLLYII 329
DB 292 DDLAAGHDVSKRLAILHYFV 312

RESULT 7
GGPP-GIBFU STANDARD; PRT; 418 AA.
AC Q92236;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Geranylgeranyl pyrophosphate synthetase (GGPP synthetase) (GGPPSASE)
DE (Geranylgeranyl diphosphate synthase) [includes:
DE Dimethylallyltransferase (EC 2.5.1.1); Geranyltransferase
DE (EC 2.5.1.10); Farnesyltransferase (EC 2.5.1.29)].

GN Name=GGPS; Synonyms=GGPPS;
OS Gibberella fujikuroi (Bakanae and foot rot disease fungus) (Fusarium moniliforme).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxID=5127;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=m567;
RG MEDLINE=97374453; PubMed=9230902;
RA Mende K., Honann V., Tudzynski B.;
RT "The geranylgeranyl diphosphate synthase gene of Gibberella fujikuroi: isolation and expression.";
RL Mol. Gen. Genet. 255:96-105(1997).
CC -!- FUNCTION: Catalyzes the trans-addition of the three molecules of IPP onto DMAPP to form geranylgeranyl pyrophosphate.
CC -!- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl diphosphate = diphosphate + geranyl diphosphate.
CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate = diphosphate + trans,trans-farnesyl diphosphate.
CC -!- CATALYTIC ACTIVITY: Trans,trans-farnesyl diphosphate + isopentenyl diphosphate = diphosphate + geranylgeranyl diphosphate.
CC -!- PATHWAY: Gibberellins biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR EMBL; CR382130; CAG81114.1; -.
SQ SEQUENCE 418 AA; 46477 MW; BE2A0F1EA7D176C9 CRC64;

Query Match 36.5%; Score 628.5; DB 1; Length 418;
Best Local Similarity 44.7%; Pred. No. 7e-36;
Matches 127; Conservative 58; Mismatches 80; Indels 19; Gaps 6;

QY 15 WSSQNESLISKPNHLLKPGKFNRLNLIQVNRVNNLPKQDLATVSOIVELLHNSLLI 74
DB 122 WTDEKENVVRGPDYVYISHPGKDFRAQLIGAFNVLDVPTSSLEVIVRVGMLHESLLI 181

QY 75 DIEDNAPLRGQTTSHLIFGVPSTINTANYMFRAMQLVSQLTTKEPLYHNLITFNEE 134
DB 182 DDVQDSSELRGFPVAHIFGVAQTINSYIVFVALQELHKNLP- - - -LITIFSDE 236

QY 135 LINLHRGQGLDIYWRDPEIPIPTQEMLYNMVNKTGGLFRILRLM-EALSFSHHGH 192
DB 237 LVNLRHGQGLDIYWRDPEIPIPTQEMLYNMVNKTGGLFRILRLM-EALSFSHHGH 291

QY 193 SLVPPFINLIGLIYQIRDDYLNKDFQMSSEKGFPAEDITEGKLSFPPIVHALNFTKQTE 252
DB 292 DCVPLVNLIGLIFQIRDDYNNLSKSKYSHNKGCEDELTGKFSFPVHSIRTNFTNLQ- - 349

QY 253 QHNEILRIILLRTSDKDIKLIQILEFPTNSLAYTKNFINOINLV 296
DB 350 - - - -LINLIKQRTSDTQIKRYAVAYME-STGSPFYTRKVLVLI 388

RESULT 8
GGPP-GIBFU STANDARD; PRT; 375 AA.
AC Q9C452;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Geranylgeranyl pyrophosphate synthetase (GGPP synthetase) (GGPPSASE)
DE (Geranylgeranyl diphosphate synthase) [includes:
DE Dimethylallyltransferase (EC 2.5.1.1); Geranyltransferase
DE (EC 2.5.1.10); Farnesyltransferase (EC 2.5.1.29)].
```

DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Geranylgeranyl pyrophosphate synthase.
 GN Name=ggs1;
 OS Penicillium paxilli.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
 RX NCBI_TaxID=70109;
 RN SEQUENCE FROM N.A.
 RP MEDLINE=21106005; PubMed=11169115;
 RA Young C., McMillan L., Telfer E., Scott B.;
 RT "Molecular cloning and genetic analysis of an indole-diterpene gene
 RT cluster from *Penicillium paxilli*."
 RL Mol. Microbiol. 39:754-764(2001).
 DR ENBL: AF279807; AAK11525.1;
 DR GO: GO:0008229; P:isoprenoid biosynthesis; IEA.
 DR InterPro: IPR000092; Polyprenyl_synth.
 DR InterPro: IPR008949; Terpenoid_synth.
 DR Pfam: PF00348; polyprenyl_synth. 1.
 DR PROSITE: PS00723; POLYPRENYL_SYNTHET_1; 1.
 DR PROSITE: PS00444; POLYPRENYL_SYNTHET_2; 1.
 DR SEQUENCE 375 AA; 41505 MW; ACA6931809643B05 CRC64;
 Query Match 36.4%; Score 625.5; DB 2; Length 375;
 Best Local Similarity 42.9%; Pred. No. 9.9e-36;
 Matches 126; Conservative 62; Mismatches 89; Indels 17; Gaps 6;
 QY 2 EAKTDELINDPWWSSQNESLISKPYNHLLKPKGNFRLNLIQVNRVNNLPKDLAIVS 61
 Db 55 KSSVDGTYKIDGTWSSKNEKILGFDYMHQHPGKDVRTQLIQAFNSGLVQPPESLAIS 114
 QY 62 QIVELHNSLLIIDIDNAPLRGQTTSLLIFGVPTINTANYFYFAMOLVSQLTKE 121
 Db 115 KVTMLHTASLLIDVDNNSVLRGVPVAVENIFGTAQINSANVYFLAQEIQOL--KN 172
 QY 122 PLVHNLITIPNEELINLHRRGGDIYWRDFELPEIIPTEQYINNMVNNKTLGLRLTLRLM 181
 Db 173 PA---AIDIVKELLNLRHGGQDLFWRDTL--TCPTDEVELEVMGNTGGLFRLAVKLM 227
 QY 182 EALSPSSHGHSLVPFNLGLIYQIRDDYLNKDFQMSSEKGFPAEDITGKLSFPPIVHA 241
 Db 228 QA---ESSTGKDCVSLNVNMGFLQICDDYLNLSNTYTHNKGLEDLTGKTSFPIHS 284
 QY 242 LNFYTKGQEQEHNELIRILLRTSDKDKLKLQILEFDTNSLAYTKNFQNL 295
 Db 285 IR-----SNPGHQLVSLKQKTDDEVKRYAVQVMQ-STGSPTHTRQVVRDL 331
 RESULT 9
 GGPP_NEUCR
 ID -GGPP_NEUCR STANDARD; PRT; 428 AA.
 AC P24322; Q7RYC0.
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Geranylgeranyl pyrophosphate synthetase (GGPP synthetase) (GGPPase)
 DE (Geranylgeranyl diphosphate synthase) (albedo-3 protein) [includes:
 DE dimethylallyltransferase (EC 2.5.1.1); Geranyltransferase
 DE (EC 2.5.1.10); Farnesyltransferase (EC 2.5.1.29)].
 GN Name=al-3; ORFNames=B8P8.010, NCU01427.1;
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 RX NCBI_TaxID=5141;
 RN SEQUENCE FROM N.A.
 RP STRAIN=74-OR23-1A / FGSC 987;
 RC MEDLINE=91170267; PubMed=1826006;
 RX Carattoli A., Romano N., Ballarín P., Morelli G., Macino G.;
 RT "The *Neurospora crassa* carotenoid biosynthetic gene (albedo 3) reveals
 RT highly conserved regions among prenilyltransferases."

RL J. Biol. Chem. 266:5854-5859(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=74-OR23-1A / FGSC 987;
 RX MEDLINE=22542210; PubMed=12655011;
 RA Manhaupt G., Montrone C., Haase D., Meves H.-W., Aign V.,
 RA Hoheisel J.D., Fartmann B., Nyakatura G., Kempken F., Maier J.,
 RA Schulte U.;
 RT "What's in the genome of a filamentous fungus? Analysis of the
 RT *Neurospora* genome sequence."
 RL Nucleic Acids Res. 31:1944-1954(2003).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=74-OR23-1A / FGSC 987;
 RX PubMed=12712197; DOI=10.1038/nature01554;
 RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
 RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
 RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
 RA Qui D., Tanakiev P., Bell-Pedersen D., Nelson M.A.,
 RA Werner-Washburne M., Selitrennikoff C.P., Kinsey J.A., Braun E.L.,
 RA Zelter A., Schulte U., Kothe G.O., Jedd G., Meves H.-W., Staben C.,
 RA Marcotte E., Greenberg D., Roy A., Foley K., Maylor J.,
 RA Stange-Thomann N., Barrett R., Gnerre S., Kamal M., Kamvysseis M.,
 RA Mauceli E., Bielke C., Rudd S., Frisman D., Krystofova S.,
 RA Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S., Cogoni C.,
 RA Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A.,
 RA Desouza C.P., Glass L., Orbach M.J., Berglund J.A., Voelker R.,
 RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
 RA Natvig D.O., Alex L.S., Manhaupt G., Bbbole D.J., Freitag M.,
 RA Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.;
 RT "The genome sequence of the filamentous fungus *Neurospora crassa*."
 RL Nature 422:859-868(2003).
 CC -!- FUNCTION: Catalyzes the trans-addition of the three molecules of
 CC IPP onto DMAPP to form geranylgeranyl pyrophosphate.
 CC -!- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl
 CC diphosphate = diphosphate + geranyl diphosphate.
 CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate
 CC = diphosphate + trans,trans-farnesyl diphosphate.
 CC -!- CATALYTIC ACTIVITY: Trans,trans-farnesyl diphosphate + isopentenyl
 CC diphosphate = diphosphate + geranylgeranyl diphosphate + isopentenyl
 CC -!- PATHWAY: Carotenoid biosynthesis.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- INDUCTION: By blue light.
 CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC ENBL: U20940; AAC13867.1; --
 CC EMBL: BX294018; CAD70868.1; --
 CC ENBL: ABE01000298; EAA31459.1; --
 CC FR: S15662; S15662.
 CC InterPro: IPR000092; Polyprenyl_synth.
 CC InterPro: IPR008949; Terpenoid_synth.
 CC Pfam: PF00348; polyprenyl_synth. 1.
 CC PROSITE: PS00723; POLYPRENYL_SYNTHET_1; 1.
 CC PROSITE: PS00444; POLYPRENYL_SYNTHET_2; 1.
 CC Carotenoid biosynthesis; Isoprene biosynthesis; Transferase.
 CC ACT SITE 268 268 By similarity.
 CC SEQUENCE 428 AA; 47887 MW; 7989DERABD8E360F CRC64;
 Query Match 36.2%; Score 623; DB 1; Length 428;
 Best Local Similarity 39.1%; Pred. No. 1.8e-35;
 Matches 125; Conservative 74; Mismatches 99; Indels 22; Gaps 7;
 QY 15 WSSQNESLISKPYNHLLKPKGNFRLNLIQVNRVNNLPKDLAIVSQTIVELHNSLLI 74
 Db 121 WSEKEKVLGTGYDYLNGHPGKDIQSVQKAPDAMLDVPSLESLEVTIKVISMULTASLLV 180

```
QY 75 DDIEDNAPLRGQTTSHLIFGVPSINTANTYMYFRAMQVLSQLTKEPLYHNLTITFEE 134
Db 181 DVEDNSVLRKGFPPVAHSIFGIPOTINTSNVYFYALQELQKL--KNP--KAVSIFSEE 235
QY 135 LINLHRGQGLDIYWRDPEIPIPTQEMYLNMVNMKGTGLFRLTLRLMEALSPSSHGHSL 194
Db 236 LINLHRGQGLDIYWRDPEIPIPTQEMYLNMVNMKGTGLFRLTLRLMEALSPSSHGHSL 290
QY 195 VPFNLGLGIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPPIVHNLFTKTKGQTHQ 254
Db 291 VPLVNIIGLIFQIADDDYHNLWREYNTKNGCEDLTEGKFSFPPIHSIR-----SNPSN 344
QY 255 NEILRIILLRTSDKDKIKLQILEFDFTNSLAYTKNFQINQVNMKNKNDENKYLPLDLASH 314
Db 345 MOLLNLIKQKGTDEEVKAYAYME-STGSFAYTRKVVREF-----DRQMTEDIDDG 398
QY 315 SDTANLHDELIIYIDHLS 334
Db 399 RKSGGIHKILDRIMLHQB 418

RESULT 10
Q6MYH1
ID Q6MYH1 PRELIMINARY; PRT; 387 AA.
AC Q6MYH1.
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE Geranylgeranyl pyrophosphate synthetase, putative (EC 2.5.1.1).
GN ORFNames=Afa63.050C;
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP SEQUENCE FROM N.A.
EX PubMed=1498527;
RA Pain A., Woodward J., Quail M.A., Anderson M.J., Clark R., Collins M.,
RA Foster N., Fraser A., Harris D., Larke N., Murphy L., Humphrey S.,
RA O'Neil S., Pettea M., Price C., Rabbinoiwtsch E., Rajadream M-A.,
RA Salzberg S., Saunders D., Seeger K., Sharp S., Warren T.,
RA Denning D.W., Barrell B., Hall N.;
RA "Insight into the genome of Aspergillus fumigatus: analysis of a 922
RT kb region encompassing the nitrate assimilation gene cluster.";
RL Fungal Genet. Biol. 41:443-453(2004).
DR EMBL; BX649606; CAF32032.1; -.
DR GO; GO:0004161; F:dimethylallyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000092; Polyprenyl synth.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00348; polyprenyl synth; 1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET_2; 1.
KW Transferase.
SQ SEQUENCE 387 AA; 43397 MW; 04F3F846C0A2A1D7 CRC64;

Query Match 35.7%; Score 614; DB 2; Length 387;
Best Local Similarity 40.8%; Pred. No. 6.6e-35;
Matches 128; Conservative 62; Mismatches 88; Indels 36; Gaps 6;

QY 12 DPWSSONESLISKPNYHLLKPKGNFRNLIVQINRNMVNMKGTGLFRLTLRLMEALSPSSHGH 71
Db 67 DGNWSEYKILMGPDYDQHQHKGIRQLITAFNMVNLQVRPESLIITKVVGLHTAS 126
QY 72 LLIDIEDNAPLRGQTTSHLIFGVPSINTANTYMYFRAMQVLSQLTKEPLYHNLTITF 131
Db 127 LLVDDVEDNSILRRGIPVAHNIFGTAQTINSANYVYFLALQEQKLNPP-----TAIDIF 181
QY 132 NEELNLHRGQGLDIYWRDPEIPIPTQEMYLNMVNMKGTGLFRLTLRLMEALSPSSHGH 191
Db 182 VQELLNLHRGQGLDIYWRDPEIPIPTQEMYLNMVNMKGTGLFRLTLRLMEALSPSSHGH 236
QY 192 HSLVFPINLLGIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPPIVHNLFTKTKGQTHQ 251
Db 237 IDCVSIVNMVGLIFQICDDYLNLSNKTYTQNGKCEDLTEGKFSFPPIHSIR-----SN 290
QY 252 EQHNEILRIILLRTSDKDKIKLQILEFDFTNSLAYTKNFQINQVNMKNKNDENKYLPLDL 311
Db 291 PRNHQISILKQRTKDEEVKALYSYME-STGSFAYTRKVVREF----- 333
QY 312 ASHSDTATNLHDEL 325
Db 334 ---RDKALSILDEI 344

RESULT 12
Q6F6E2
ID Q6F6E2 PRELIMINARY; PRT; 343 AA.
AC Q6F6E2;
```

```
QY 192 HSLVFPINLLGIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPPIVHNLFTKTKGQTHQ 251
Db 237 IDCVSIVNMVGLIFQICDDYLNLSNKTYTQNGKCEDLTEGKFSFPPIHSIR-----SN 290
QY 252 EQHNEILRIILLRTSDKDKIKLQILEFDFTNSLAYTKNFQINQVNMKNKNDENKYLPLDL 311
Db 291 PRNHQISILKQRTKDEEVKALYSYME-STGSFAYTRKVVREF----- 333
QY 312 ASHSDTATNLHDEL 325
Db 334 ---RDKALSILDEI 344

RESULT 11
CAF32032
ID CAF32032 PRELIMINARY; PRT; 387 AA.
AC CAF32032;
DT 13-MAY-2004 (TREMBlrel. 27, Created)
DT 13-MAY-2004 (TREMBlrel. 27, Last sequence update)
DT 13-MAY-2004 (TREMBlrel. 27, Last annotation update)
DE Geranylgeranyl pyrophosphate synthetase, putative (EC 2.5.1.1).
GN Afa6E3.050C;
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP SEQUENCE FROM N.A.
RA Pain A., Woodward J., Quail M.A., Anderson M.J., Clark R., Collins M.,
RA Foster N., Fraser A., Harris D., Larke N., Murphy L., Humphrey S.,
RA O'Neil S., Pettea M., Price C., Rabbinoiwtsch E., Rajadream M-A.,
RA Salzberg S., Saunders D., Seeger K., Sharp S., Warren T.,
RA Denning D.W., Barrell B., Hall N.;
RA "Insight into the genome of Aspergillus fumigatus: analysis of a 922
RT kb region encompassing the nitrate assimilation gene cluster.";
RL Fungal Genet. Biol. 41:443-453(2004).
DR EMBL; BX649606; CAF32032.1; -.
DR EMBL; BX649606; CAF32032.1; -.
KW Transferase.
SQ SEQUENCE 387 AA; 43397 MW; 04F3F846C0A2A1D7 CRC64;

Query Match 35.7%; Score 614; DB 2; Length 387;
Best Local Similarity 40.8%; Pred. No. 6.6e-35;
Matches 128; Conservative 62; Mismatches 88; Indels 36; Gaps 6;

QY 12 DPWSSONESLISKPNYHLLKPKGNFRNLIVQINRNMVNMKGTGLFRLTLRLMEALSPSSHGH 71
Db 67 DGNWSEYKILMGPDYDQHQHKGIRQLITAFNMVNLQVRPESLIITKVVGLHTAS 126
QY 72 LLIDIEDNAPLRGQTTSHLIFGVPSINTANTYMYFRAMQVLSQLTKEPLYHNLTITF 131
Db 127 LLVDDVEDNSILRRGIPVAHNIFGTAQTINSANYVYFLALQEQKLNPP-----TAIDIF 181
QY 132 NEELNLHRGQGLDIYWRDPEIPIPTQEMYLNMVNMKGTGLFRLTLRLMEALSPSSHGH 191
Db 182 VQELLNLHRGQGLDIYWRDPEIPIPTQEMYLNMVNMKGTGLFRLTLRLMEALSPSSHGH 236
QY 192 HSLVFPINLLGIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPPIVHNLFTKTKGQTHQ 251
Db 237 IDCVSIVNMVGLIFQICDDYLNLSNKTYTQNGKCEDLTEGKFSFPPIHSIR-----SN 290
QY 252 EQHNEILRIILLRTSDKDKIKLQILEFDFTNSLAYTKNFQINQVNMKNKNDENKYLPLDL 311
Db 291 PRNHQISILKQRTKDEEVKALYSYME-STGSFAYTRKVVREF----- 333
QY 312 ASHSDTATNLHDEL 325
Db 334 ---RDKALSILDEI 344

RESULT 12
Q6F6E2
ID Q6F6E2 PRELIMINARY; PRT; 343 AA.
AC Q6F6E2;
```

```

DT 01-OCT-2004 (TREMBlrel. 28, Created)
DT 01-OCT-2004 (TREMBlrel. 28, Last sequence update)
DT 01-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE Geranylgeranyl diphosphate synthase.
GN Name=PbGGS;
OS Phoma betae.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Phoma.
OX NCBI_TaxID=137527;
RN [1]
RP SEQUENCE FROM N.A.
RA Toyomasu T., Nakaminami K., Toshima H., Mie T., Watanabe K., Ito H.,
RA Matsui H., Mitsuhashi W., Sassa T., Oikawa H.;
RT "Cloning of a Gene Cluster Responsible for the Biosynthesis of
RT Diterpene Aphidicolin, a Specific Inhibitor of DNA Polymerase alpha.";
RL Biosci. Biotechnol. Biochem. 68:146-152(2004).
DR EMBL; AB079897; BAD29965.1; -
SQ SEQUENCE 343 AA; 38975 MW; 5C873525B973CB50 CRC64;

Query Match 35.6%; Score 612.5; DB 2; Length 343;
Best Local Similarity 43.0%; Pred. No. 7.2e-35;
Matches 123; Conservative 60; Mismatches 86; Indels 17; Gaps 6;

QY 15 WSSQNESLISPKYNHLLKPKGNFRNLIVQINRVNMLPKDLAIVSQIVVELLHNSLLI 74
DB 48 WTSTKEKVVTPGYDYIAAKPKGEVRTLACDFDEWLQVPPSPLESLEVGQVVRMLHTASLLI 107

QY 75 DDIEDNAPLRGQTTSHLIFGVPTINTANTANYFRAMQLVSQLTTKEPLYHNLITIFNEE 134
DB 108 DDIQNSLRGKPKVAQNIQFALTINSANYVFLALEKINSL--KNP---NITDIFTEE 162

QY 135 LINHRGGLDIYWRDFLPEIPIQEMLYNMVNKTKGLFRLTLRLMEALSPSSHGHSL 194
DB 163 LLRLHRGQAMLDLYWDTL--TCPTEEYFEWVANVTGGLFWLMYRMMA--ESSMPIDL 217

QY 195 VPFNLGLIYQIRDDYLNKDFQMSSEKGAEDITEKLSFPPIVHALNFTKTKGQTOHQ 254
DB 218 LPVVELLGVIFQVDDYKNCRSRYGKLGKGFEDLTEGKFSFPVHSIR-----SNPED 271

QY 255 NEILRIILLRTSDKDIKLKQLILEFDNLSLAYTKNFQINQVNMK 300
DB 272 LQLHLVQQKSSNEHVKLYAIEIME-STGSLEYTKHVVENIVSQIQ 316

Query Match 34.8%; Score 598.5; DB 2; Length 343;
Best Local Similarity 42.3%; Pred. No. 6.9e-34;
Matches 121; Conservative 59; Mismatches 89; Indels 17; Gaps 6;

QY 15 WSSQNESLISPKYNHLLKPKGNFRNLIVQINRVNMLPKDLAIVSQIVVELLHNSLLI 74
DB 48 WTSTKEKVVTPGYDYIAAKPKGEVRTLACDFDEWLQVPPSPLESLEVGQVVRMLHTASLLI 107

```

```

QY 75 DDIEDNAPLRGQTTSHLIFGVPTINTANTANYFRAMQLVSQLTTKEPLYHNLITIFNEE 134
DB 108 DDIQNSLRGKPKVAQNIQFALTINSANYVFLALEKINSL--KNP---NITDIFTEE 162
QY 135 LINHRGGLDIYWRDFLPEIPIQEMLYNMVNKTKGLFRLTLRLMEALSPSSHGHSL 194
DB 163 LLRLHRGQAMLDLYWDTL--TCPTEEYFEWVANVTGGLFWLMYRMMA--ESSMPIDL 217
QY 195 VPFNLGLIYQIRDDYLNKDFQMSSEKGAEDITEKLSFPPIVHALNFTKTKGQTOHQ 254
DB 218 LPVVELLGVIFQVDDYKNCRSRYGKLGKGFEDLTEGKFSFPVHSIR-----SNPED 271
QY 255 NEILRIILLRTSDKDIKLKQLILEFDNLSLAYTKNFQINQVNMK 300
DB 272 LQLHLVQQKSSNEHVKLYAIEIME-STGSLEYTKHVVENIVSQIQ 316

RESULT 14
Q9P885
ID Q9P885 PRELIMINARY; PRT; 303 AA.
AC Q9P885;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Geranylgeranyl pyrophosphate synthase (EC 2.5.1.29).
GN Name=carg;
OS Mucor circinelloides f. lusitanicus.
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OX Mucor.
OC NCBI_TaxID=29924;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS277.49;
RA MEDLINE=22581058; PubMed=12695851;
RA Velayos A., Papp T., Aguilar-Elena R., Fuentes-Vicente M.,
RA Esalva A.P., Iturriga E.A., Alvarez M.I.;
RT "Expression of the carg gene, encoding geranylgeranyl pyrophosphate
RT synthase, is up-regulated by blue light in Mucor circinelloides.";
RT Curr. Genet. 43:112-120(2003).
DR EMBL; AJ276129; CAB89115.1; -.
DR GO; GO:0004311; F:farnesyltransferase activity; IEA.
DR GO; GO:0016740; F:transferrase activity; IEA.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyprenyl_synth.
DR InterPro; IPR008949; Terpenoid_synth.
DR Pfam; PF00348; polyprenyl_synth; 1.
DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; UNKNOWN_1.
DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
DR Transferase.
SQ SEQUENCE 303 AA; 34787 MW; D440757CA7D0E215 CRC64;

Query Match 34.4%; Score 591.5; DB 2; Length 303;
Best Local Similarity 41.1%; Pred. No. 1.8e-33;
Matches 113; Conservative 67; Mismatches 78; Indels 17; Gaps 5;

QY 20 ESLISPKYNHLLKPKGNFRNLIVQINRVNMLPKDLAIVSQIVVELLHNSLLIDIED 79
DB 14 EDILEPYYLLISQPKDIAKLIAPDLWLHVPKDVLGVINKIIGHLNHLASLMDVQD 73

QY 80 NAPLRGQTTSHLIFGVPTINTANTANYFRAMQLVSQLTTKEPLYHNLITIFNEELNLH 139
DB 74 DSDLRGVPAHLYGVFPQINTANYVIFLALQEVWMLNIP-----SNMQVCTEELNLH 128

QY 140 RGQGLDIYWRDFLPEIPIQEMLYNMVNKTKGLFRLTLRLMEALSPSSHGHSLVFPIN 199
DB 129 RGQGLIYWRDSL--TCPTEEYIDVMVNKTSGLLRLAVALMQAASD---IDYTLVN 183

QY 200 LGLIYQIRDDYLNKDFQMSSEKGAEDITEKLSFPPIVHALNFTKTKGQTOHQ 259
DB 184 IIGHFQVRDDYVNMQLQSTSYNNKGFCEDLTEGKFSFPVHSIR-----KDPNSRLN 237

QY 260 ILLRTSDKDIKLKQLILEFDNLSLAYTKNFQINQ 294

```

Db 238 IISQPTSIKVKYALEVIR-KAGSFYRFLRQ 271

RESULT 15

GGPP_HUMAN STANDARD; PRT; 300 AA.

AC O95749;

AD 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 01-OCT-2004 (Rel. 45, Last annotation update)

DE Geranylgeranyl pyrophosphate synthetase (GGPP synthetase) (GGPPSASE)

DE (Geranylgeranyl diphosphate synthase) [Includes:

DE Dimethylallyltransferase (EC 2.5.1.1); Geranyltransferase

DE (EC 2.5.1.10); Farnesyltransferase (EC 2.5.1.29)].

GN Name=GGPS1;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal heart;

RK MEDLINE=98412715; PubMed=9741684;

RA Ericsson J., Greene J.M., Carter K.C., Shell B.K., Duan D.R.,

RA Florence C., Edwards P.A.;

RT "Human geranylgeranyl diphosphate synthase: isolation of the cDNA,

RT chromosomal mapping and tissue expression.";

RL J. Lipid Res. 39:1731-1739(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RK MEDLINE=99150380; PubMed=10026212;

RA Kuzuguchi T., Morita Y., Segami I., Segami H., Ogura K.;

RT "Human geranylgeranyl diphosphate synthase. cDNA cloning and

RT expression.";

RL J. Biol. Chem. 274:5888-5894(1999).

RN [3]

RP SEQUENCE FROM N.A.

RC Misawa N., Okazaki H., Noguchi Y., Tatsuno I., Saito Y., Yasuda T.,

RA Hirai A.;

RT "Study on isolation of a geranylgeranyl pyrophosphate (GGPP) synthase

RT cDNA and its expression - development of a new assay system of gene

RT functions.";

RL Proc. Jpn. Conf. Biochem. Lipids 41:293-296(1999).

RN [4]

RP SEQUENCE FROM N.A.

RC TISSUE=Epituitary;

RK MEDLINE=20402571; PubMed=10931946;

RA Hu R.-M., Han Z.-G., Song H.-D., Peng Y.-D., Huang Q.-H., Ren S.-X.,

RA Gu Y.-J., Huang C.-H., Li Y.-B., Jiang C.-L., Fu G., Zhang Q.-H.,

RA Gu B.-W., Dai N., Mao Y.-F., Gao G.-F., Rong R., Ye M., Zhou J.,

RA Xu S.-H., Gu J., Shi J.-X., Jin W.-R., Zhang C.-K., Wu T.-M.,

RA Huang G.-Y., Chen Z., Chen M.-D., Chen J.-L.;

RT "Gene expression profiling in the human hypothalamus-pituitary-adrenal

RT axis and full-length cDNA cloning.";

RL Proc. Natl. Acad. Sci. U.S.A. 97:9543-9548(2000).

RN [5]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver, and Spleen;

RK MEDLINE=99203156; PubMed=10101267;

RA Kainou T., Kawamura K., Tanaka K., Matsuda H., Kawamukai M.;

RT "Identification of the GGPS1 genes encoding geranylgeranyl diphosphate

RT synthases from mouse and human.";

RL Biochim. Biophys. Acta 1437:333-340(1999).

RN [6]

RP SEQUENCE FROM N.A.

RC Zhang M., Yu L., Hu P., Bi A., Zhang Q., Xu M., Zhao S.;

RT "Molecular cloning and expression analysis of a novel human cDNA

RT encoding a protein homologous to Neurospora crassa geranylgeranyl

RT pyrophosphate synthetase.";

RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

RN [7]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RK MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo A.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.J., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -!- FUNCTION: Catalyzes the trans-addition of the three molecules of

CC IPP onto DMAPP to form geranylgeranyl pyrophosphate, an important

CC precursor of carotenoids and geranylated proteins.

CC -!- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl

CC diphosphate = diphosphate + geranyl diphosphate.

CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate

CC = diphosphate + trans,trans-farnesyl diphosphate.

CC -!- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl

CC diphosphate = diphosphate + geranylgeranyl diphosphate.

CC -!- PATHWAY: Isoprenoid biosynthesis.

CC -!- SUBUNIT: Homocotamer.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- TISSUE SPECIFICITY: Abundantly expressed in testis. Found in other

CC tissues to a lower extent.

CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; AB017971; BAA75909.1; -

CC EMBL; AB019036; BAA77251.1; -

CC EMBL; AF125394; AAD43050.1; -

CC EMBL; AB016043; BAA76511.1; -

CC EMBL; AF057698; AAG45581.1; -

CC EMBL; BC005252; AAH05252.1; -

CC Genew; HGNC:4249; GGPS1.

CC MIM; 606982; -

CC GO; GO:0004337; F:geranyltransferase activity; TAS.

CC GO; GO:0006629; P:lipid metabolism; TAS.

CC InterPro; IPR000092; Polyprenyl synth.

CC InterPro; IPR008949; Terpenoid synth.

CC Pfam; PF00348; polyprenyl_synth; 1.

CC PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.

CC PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.

CC Isoprene biosynthesis; Transferase.

CC KW SEQUENCE 300 AA; 34871 MW; F5D1959274BEE27A CRC64;

CC

Query Match 33.9%; Score 583; DB 1; Length 300;

Best Local Similarity 40.5%; Pred. No. 7.1e-33;

Matches 128; Conservative 55; Mismatches 103; Indels 30; Gaps 8;

QY 20 ESLISPYNHILKPGKFNFLNLVQINRVMNLPKDLAIQSVQIVELLNSSLIDIED 79

Db 9 QRILLEPYKLLQLPGKQVTKLSQAFNHLKVPDKLQIIIEVTMLNHLASLLIDIED 68

QY 80 NAPLRGQTTSHLIFGVPSTINTANTMYFRAMQIVSQLTTPKEPLYHNLITFNEELNLH 139

Db	69	NSKLRGPFVAHSIYGIPSVINSANYVFLGKVV--LTIIDHP--DAVKLFTRQLLEIH	123
Qy	140	RGQGLDIYWRDFLPBIIPTQEMYLNMVNMKTGGLFELTLRLMEALSPSSHGHGHSIVPFIN	199
Db	124	QOQGLDIYWRD--NYTCPEEYKAMVLOKTGGLFGLAVGLMQLP---SDYKEDLKPLIN	178
Qy	200	LLGIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPPIVHALNFTKTKGQTEQHNEILR	259
Db	179	TLGLFFQIRDDYANLHSEKSEKSFCELTGKFSFPTIHAI-----WSRPESTQVQN	232
Qy	260	ILLRTSDKDIKLIQILEFTNSLAYTKNFINOIVNMKNENKYLFDLASHSDTAT	319
Db	233	ILRQFTENIDIKKYCVHYLE-DVGSFEYTRNTLKL-----EAK-----AYKQIDAR	278
Qy	320	NLHDELLYIIDHLSL	335
Db	279	GGNPELVALVKHLSKM	294

Search completed: October 17, 2004, 08:52:57
Job time : 125.839 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 08:05:06 ; Search time 21.8543 Seconds
(without alignments)
1474.884 Million cell updates/sec

Title: US-10-041-018-22

Perfect score: 1720

Sequence: 1 MEAKIDELINNDPVMSSQNE.....DTATNLHDELIIIDHSEL 335

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1720	100.0	335	2 S60921	farnesyltranstrans
2	623	36.2	428	2 S15662	farnesyltranstrans
3	337	19.6	347	2 A70139	octaprenyl-diphosp
4	310.5	18.1	327	2 D64407	bifunctional short
5	274.5	16.0	364	2 C71295	probable octapreny
6	244	14.2	325	2 T45152	bifunctional short
7	243.5	14.2	323	2 AB0427	octaprenyl-diphosp
8	239	13.9	325	2 G69165	bifunctional short
9	238.5	13.9	324	2 A81213	octaprenyl-diphosp
10	233	13.5	322	2 C83075	octaprenyl-diphosp
11	228.5	13.3	330	2 A54058	octaprenyl-diphosp
12	226	13.1	323	2 AH0903	farnesyltranstrans
13	223	13.0	323	2 E85982	octaprenyl-diphosp
14	223	13.0	323	2 B91137	octaprenyl diphosp
15	223	13.0	323	2 B65109	octaprenyl-diphosp
16	214.5	12.5	332	2 A82688	octaprenyl-diphosp
17	211.5	12.3	332	2 H90145	hypothetical prote
18	210.5	12.2	330	2 A87194	polyprenyl diphosp
19	210.5	12.2	332	2 S75427	farnesyltranstrans
20	209.5	12.2	346	2 C75139	bifunctional short
21	205	11.9	323	2 A40433	prephytoene pyroph
22	205	11.9	323	2 T06874	probable prenyl tr
23	204	11.9	329	2 L64160	hypothetical prote
24	203	11.8	297	2 B81400	polyprenyl synthet
25	202	11.7	348	2 A82322	octaprenyl-diphosp
26	201	11.7	299	2 C72242	hypothetical prote
27	198	11.5	342	2 E71101	probable geranylge
28	197.5	11.5	338	2 H87510	polyprenyl synthet
29	195.5	11.4	304	2 A70377	octoprenyl-diphosp

ALIGNMENTS

RESULT 1

S60921

farnesyltranstransferase (EC 2.5.1.29) - yeast (Saccharomyces cerevisiae)

N:Alternate names: geranylgeranyl-diphosphate synthase; protein lPE1c; protein YPL069c

C:Species: Saccharomyces cerevisiae

C>Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004

C/Accession: S60921; S61936

R:Winnett, E.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Hall, J.; Storms, R.K.;

submitted to the EMBL Data Library, October 1995

A:Reference number: S60921

A:Accession: S60921

A:Molecule type: DNA

A:Residues: 1-335 <WIN>

A:Cross-references: UNIPROT:Q12051; EMBL:U39205; NID:gi079672; PIDN:AAB68286.1; PID:gl079672;

R:Jiang, Y.; Proteau, P.; Poulter, D.; Ferro-Novick, S.

submitted to the EMBL Data Library, July 1995

A:Description: Bts1 encodes a geranylgeranyl diphosphate synthase in Saccharomyces cerevisiae

A:Reference number: S61936

A:Accession: S61936

A:Molecule type: DNA

A:Residues: 1-335 <JIA>

A:Cross-references: EMBL:U31632; NID:gl098640; PIDN:AAA83262.1; PID:gl098641

C:Genetics:

A:Gene: SGD:BTS1

A:Cross-references: SGD:S0005990; MIPS:YPL069c

A:Map position: 16L

C:Superfamily: prenyl transferase A

C:Keywords: transferase

Query Match 100.0%; Score 1720; DB 2; Length 335;
Best Local Similarity 100.0%; Pred. No. 4.3e-114;
Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MEAKIDELINNDPVMSSQNESLSKPNHLLKPGKFRNLIVQINRVNVLPRDQLAIV	60
Db	1	MEAKIDELINNDPVMSSQNESLSKPNHLLKPGKFRNLIVQINRVNVLPRDQLAIV	60
QY	61	SQIVELHNSLLIDIEDNAPLRGQTTSHLIFGVPSITNTANTMYFRAMQLVSLTTK	120
Db	61	SQIVELHNSLLIDIEDNAPLRGQTTSHLIFGVPSITNTANTMYFRAMQLVSLTTK	120
QY	121	EPIYHNLIITFNEELINLHRGQGLDIYWRDFFPEIIPTOEYLNVMVNTGGFLTRLRL	180
Db	121	EPIYHNLIITFNEELINLHRGQGLDIYWRDFFPEIIPTOEYLNVMVNTGGFLTRLRL	180
QY	181	MEALSPSSHGHSHSLVPFINLGIYQIRDDYLNLDKDFQMSSEKGFADITEGKLSFPIVH	240
Db	181	MEALSPSSHGHSHSLVPFINLGIYQIRDDYLNLDKDFQMSSEKGFADITEGKLSFPIVH	240
QY	241	ALNFTKTKGQTEQHNEILRLITSDKDKIKLQILEFDNTSLAYTKNFNLQNLVNMK	300
Db	241	ALNFTKTKGQTEQHNEILRLITSDKDKIKLQILEFDNTSLAYTKNFNLQNLVNMK	300

octaprenyl-diphosphate synthase (ispB) homolog - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1996 #text_change 09-Jul-2004
C:Accession: A70139
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, son, D.; Peterson, J.; Kurlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, i.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B. Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: A70139
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-347 <KLE>
A:Cross-references: UNIPROT:O51294; GB:AE001138; GB:AE000783; NID:G2688210; PIDN:AAC56699
A:Experimental source: strain B31
C:Superfamily: prenyl transferase A

```

QY      297 NMIKNDNE 304
      : | :
Db      338 DTIEQIK 345

RESULT 4
D64407
Bifunctional short chain isoprenyl diphosphate synthase homolog - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C:Accession: D64407
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.J.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodex, A.;
ison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A>Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8689087
A:Accession: D64407
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-327 <BUL>
A:Cross-references: UNIPROT:Q58270; GB:U67530; GB:L77117; NID:G2826340; PIDN:AA898865.1
C:Genetics:
A:Map position: REV783844-782861
A:Start codon: TTG

```

Query Match	18.1%;	Score 310.5;	DB 2;	Length 327;
Best Local Similarity	27.6%;	Pred. No. 1.2e-14;		

RESULT 6
T45152
Bifunctional short chain isoprenyl diphosphate synthase [imported] - Methanobacterium thurstonii
C:Species: Methanobacterium thermoautotrophicum
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
C:Accession: T45152
R:Chen, A.; Poulter, C.D.
Arch. Biochem. Biophys. 314, 399-404, 1994
A:Title: Isolation and characterization of *idsA*: the gene for the short chain isoprenyl diphosphate synthase from *Methanobacterium thermoautotrophicum*
A:Reference number: Z22932; MUID:95070145; PMID:7979381
A:Accession: T45152
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-325 <CHE>
A:Cross-references: EMBL:S75695; PIDN:AAB32421.1
C:Genetics:
A:Gene: *idsA*
C:Superfamily: prenyl transferase A

Query Match 14.2%; Score 244; DB 2; Length 325;
Best Local Similarity 26.3%; Pred. No. 5.6e-10;
Matches 82; Conservative 61; Mismatches 145; Indels 24; Gaps 10;

QY 25 KPNHILKPGKNFRILN-IVQINRVMLPKDQLAIVSOIVELHNSLILDDIEDNAPL 83
DB 33 KASEHLITAGGKKIRPSLALLSCAEVGNPED-AGVAAAEHLITFSLIHDDINDDEM 91
QY 84 RRGQTTSHLIFGVPTINTANVYFRAMQLVSQLTKEPLVHNLITIFNEELINLHRCGG 143
DB 92 RRGEPSPVHWIGEPMAILAGDVLFSKAFPAVIRNGDSERV-KDALAVVVDSCVKCEGQA 150
QY 144 LDYWRDELPEIIPTEQMYLNNVMVNMKTGGLFLTLRLMEALSPSSHG-HSLVPPFINLLG 202
DB 151 LDM---GFEERLDVTEDETMNIYKKTAAIAATKAGAIMGCASEREVEALEDYKFTG 207
QY 203 ITYQIRDDYLNILKDFQMSSEKGFARDITEGKLSPFIVHALNFTKTKGOTECHNEILRLIL 262
DB 208 LAFQIHDDYLDVVSDSEISLGKPVGSDIAEGKWTLMVVAL-----EEASEEDREL-LSI 261
QY 263 LRTSKDKLKLQILEPDTSLATYKNFNLQNLVNMKNENKYLPLDASHSDPATNLH 322
DB 262 LGSDEGSAEAEIETFE-RYGATQYAEHVADYVYRMAKERLE-----ILESDA---R 310
QY 323 DELLVIIDHLSL 334
DB 311 DALMRIADEVLE 322

RESULT 7
AB0427
octaprenyl-diphosphate synthase (EC 2.5.1.-) [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, S.;
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AB0427
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-323 <KUP>
A:Cross-references: UNIPROT:Q8ZEA5; GB:AL590842; PIDN:CAC92742.1; PID:gi5981437; GSPDB:G
C:Genetics:
A:Gene: *ispB*
C:Superfamily: prenyl transferase A
C:Keywords: transferase

Query Match 14.2%; Score 243.5; DB 2; Length 323;
Best Local Similarity 24.8%; Pred. No. 6e-10;

A;Cross-references: UNIPROT:Q9HVL5; GB:AE004870; NID:9950807; PIDN:AAG0795
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: ispB; PA4569
C;Superfamily: prenyl transferase A

Query Match	13.5%; Score 233; DB 2; Length 322;
Best Local Similarity	21.5%; Pred. No. 3.3e-09;
Matches	65; Conservative 81; Mismatches 135; Indels 22; Gaps 6;
QY	22 LIKPYNHLLKCKENFRNLIVQINRYMMLPKDQLAIVSQIVELLHNSSLIDIDEDNA 81
Db	31 LVEKIGDYIISAGGRULPVLVLAGTGLGYKGDCLCLAATIEFLTSLIHDDVDVDA 90
QY	82 PLRRGQTTSHLIFGVPSTINTANYMFRMLQVLSQLTKEPLYNHLITITNEELINLHRG 141
Db	91 GLRGRSTANALWGNARSVLGVDFLYARSEFMVVELGMP-----VMRIISQATRVIAEG 145
QY	142 QGLDIYWRDFLPEIIPFQEMYLNVMNMKTKGLFRLTLRLMEAL-SPSSHHGSHSVFPINL 200
Db	146 EVLQL-----SKVRDASTTEETMYEVIRKTAMLFEASTHSAALCQAGEEQSEALRREGDY 202
QY	201 LGIYQIRDDYINLKDFQMSSEKGFADITEGKLSPPIVHALNFTKYGQTEQHNEILRI 260
Db	203 LGIAFQVLDLLYRGDAATLGRNVGDDLAEGKPTLPLIVTM-----RDGTEEQAAALVRK 257
QY	261 LLIRTSDKDIK-----LKLILQILEEDFNSLAYTKNFINQLVNMKNENKYLPLDIASHS 315
Db	258 AIOQGSQDLESVCAAVEAAGALDYTAN-----LARDYAARAIACTLDLPDNEYRSALVELS 314
QY	316 DTA 318
Db	315 EFA 317

RESULT 11
AS4058
farnesyltransferase (BC 2.5.1.29) - Sulfolobus acidocaldarius
N:Alternate names: geranylgeranyl-diphosphate synthase
C:Species: Sulfolobus acidocaldarius
C:Date: 02-Aug-1994 #sequence_revision 02-Aug-1994 #text_change 09-Jul-2004
C:Accession: A54058
R:Ohnuma, S.; Suzuki, M.; Nishino, T.
J. Biol. Chem. 269, 14792-14797, 1994
A:Title: Archaeobacterial ether-linked lipid biosynthetic gene. Expression cloning, sequence
A:Reference number: AS4058; MUID:94237902; PMID:8182085
A:Accession: A54058
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-330 <OHN>
A:Cross-references: UNIPROT:P39464; GB:D28748; NID:G506370; PID:G506371
C:Superfamily: prenyl transferase A
C:Keywords: transferase

Query Match	13.3%;	Score 228.5;	DB 2;	Length 330;
Best Local Similarity	25.8%;	Pred. No. 7.1e-09;		
Matches	83;	Conservative 65;	Mismatches 151;	Indels 23; Gaps 10;
QY	6	DELINN-DPWSQNESLSIKPN--HILLKGNFRMLIVQINRMVNLPKDQLAIVS	61	
dbb	9	NEIVNSVNDLIKSYIGDVPKLYEASVHLTSGGKRLRPLILITSSPLFGQERAYYAG	68	
QY	62	QIVVELLHNSLLDIDDNAPLRGQQTSLHIEGVPTINTANYVPRAMQLVSQLTATKE	121	
dbb	69	AAAEVLVHTFTLVHDDIMDQDNIRGRLPTVHVYGLPLAILAGDLLHAKAQQLLTQALRGL	128	
QY	122	P--LVYHNLTIFNEEELINLHRGGDLIYWDPELPIPTQBMVYLNMYNKTGGLFRUTLR	179	
dbb	129	PSETIIKAPDIFRSLIIISGGVAMFEEDRID--IKQOE-YLDMRSRTKTAALFSASSS	185	
QY	180	LMELASPSHHGSLV-PTFNLGLIYQIRDDYINLKDQFMSSEKGFPAEDITECKLSFPI	238	
dbb	186	IGALLAGANDNDVRLMSDFQTNLGIATQIVDDTLGLTADSKELGKPFVSDIRGSKTIV	245	

[illegible]

Query Match	13.1%	Score 226;	DB 2;	Length 323;
Best Local Similarity	24.3%;	Pred. No. 1e-08;		
Matches	72;	Conservative	75;	Mismatches 113;
				Indels 38;
				Gaps 12;
QY	1	MEAKIDELINNDPVMWSQNESLISKPNYHILLKPKGNFRNLIVQINRWMLPKDQIAIV	60	
DB	17	VNATILEQLNSD-----VQLINQLGYIISGGKIRPMIAVLAARAVGYQGNHVTI	69	
QY	61	SQVLEILLNSSLLIDIEDNAPLRGOTTSHLIFGVPESTINTANYMFRAMOLVSQITTK	120	
DB	70	AALIEFTIHTALLHDVDESDMRGKATANAFCGNAASVLGVDFIYTRAFQMMTSLGSL	129	
QY	121	EPLYHNLITIFNEELINLHRGQGLDIY-WRDFPELIIPTOEMYLNMWNKTTGGLFRLTLR	179	
DB	130	K-----VLEWMSAVNVIAEGEVLQMNVD--PDI--TEENTWRVIYKTARLFEEAAQ	180	
QY	180	---LMEALSPSSHGHSLVPFFINLLGIYQIRDYLNLKDFQMSSE---KGFADITEGK	233	
DB	181	CSGILAGCTPEQKQ--LQDYGRYLGTAFLQIDDLL--DYSADGELHGNVGDLDNEGK	235	
QY	234	LSPIIVHALNFTKTKQCTQHNEILLRLTTSDDKIKLKIQLPEPT--NSLAYTK	289	
DB	236	PTLPLLHAM-----RGFTPEQSAMITAEQNGRHL---LEPVLAMITCGSLEWTR	285	

RESULT 13
E85982
octaprenyl diphosphate synthase [imported] - Escherichia coli (strain O157:H7, substrain C)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: E85982
R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85982
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-323 <STO>

A;Cross-references: UNIPROT:Q8X9K0; GB:AE005174; NID:G12517802; PIDN:AG58321.1; GSPDB:G
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A:Gene: ispB
C:Superfamily: prenyl transferase A

Query Match 13.0%; Score 223; DB 2; Length 323;
Best Local Similarity 25.4%; Pred. No. 1.7e-08;
Matches 66; Conservative 63; Mismatches 99; Indels 32; Gaps 10;

QY 1 MEAKIDELINNDPVWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVWNLPKDQLAIV 60
Db 17 VNAALIEQLNSD-----VQLINQLGYIVSGGKRIRPMIAVLAARAVGEGNAHVIT 69

QY 61 SQIVELLHNSLLIDIEDNAPLRGQTTSHLIFGVPSITINTANYMYFRAMOLVSLQTTK 120
Db 70 AALIEFIHTATLLHDDVDVDESDMRRGKATANAAGNAASVLVGDFTYTRAFQMTSLGSL 129

QY 121 EPLYHNLTITFNEELINLHRGQGLDIY-WRDFLPEIPIPTQEMVYLNVMNKTGGLFELTLR 179
Db 130 K-----VLEWSEAVNVIAEGEVLQLMNVND--PDI--TEENYMRVIYKTAFLFEAAQ 180

QY 180 ---LMEALSPSSHGHSHLVPFNLGLIYQIRDDYLNLDKDFQMSSE---KGFAEDITEGK 233
Db 181 CSGIAGCTPEEKG--LDYGRYLGTAFLQIDLL--DYNADGEQLGNVGDLLNEGK 235

QY 234 LSPPIVHALNFTKGTQTEQ 253
Db 236 PTLPLLLHAMH-----HGTPEQ 251

Search completed: October 17, 2004, 08:54:19
Job time : 21.8543 secs

Query Match 13.0%; Score 223; DB 2; Length 323;
Best Local Similarity 25.4%; Pred. No. 1.7e-08;
Matches 66; Conservative 63; Mismatches 99; Indels 32; Gaps 10;

QY 1 MEAKIDELINNDPVWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVWNLPKDQLAIV 60
Db 17 VNAALIEQLNSD-----VQLINQLGYIVSGGKRIRPMIAVLAARAVGEGNAHVIT 69

QY 61 SQIVELLHNSLLIDIEDNAPLRGQTTSHLIFGVPSITINTANYMYFRAMOLVSLQTTK 120
Db 70 AALIEFIHTATLLHDDVDVDESDMRRGKATANAAGNAASVLVGDFTYTRAFQMTSLGSL 129

QY 121 EPLYHNLTITFNEELINLHRGQGLDIY-WRDFLPEIPIPTQEMVYLNVMNKTGGLFELTLR 179
Db 130 K-----VLEWSEAVNVIAEGEVLQLMNVND--PDI--TEENYMRVIYKTAFLFEAAQ 180

QY 180 ---LMEALSPSSHGHSHLVPFNLGLIYQIRDDYLNLDKDFQMSSE---KGFAEDITEGK 233
Db 181 CSGIAGCTPEEKG--LDYGRYLGTAFLQIDLL--DYNADGEQLGNVGDLLNEGK 235

QY 234 LSPPIVHALNFTKGTQTEQ 253
Db 236 PTLPLLLHAMH-----HGTPEQ 251

Db 236 PTLPLLLHAMH-----HGTPEQ 251

RESULT 15
E65109
octaprenyl-diphosphate synthase (EC 2.5.1.-) - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
R:Accession: E65109; PV0010
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
.A.; Rose, D.J.; Mau, B.; Shaoh, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: E65109
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-323 <BLAT>
A:Cross-references: UNIPROT:P19641; GB:AE000399; GB:U00096; NID:G2367201; PIDN:AACT6219.
A:Experimental source: strain K-12, substrain MG1655
R:Choi, Y.L.; Nishida, T.; Kawamukai, M.; Utsumi, R.; Sakai, H.; Komano, T.
J. Bacteriol. 171, 5222-5225, 1989
A:Title: Cloning and sequencing of an Escherichia coli gene, nlp, highly homologous to
A:Reference number: J00027; MUID:89359178; PMID:2670911
A:Accession: PV0010
A:Molecule type: DNA
A:Residues: 1, 156-157, 196-323 <CHO>
A:Cross-references: GB:X68873; NID:G42128; PIDN:CAA48735.1; PID:G388220
C:Genetics:
A:Gene: ispB
A:Map position: 69 min
C:Superfamily: prenyl transferase A
C:Keywords: transferase

Query Match 13.0%; Score 223; DB 2; Length 323;
Best Local Similarity 25.4%; Pred. No. 1.7e-08;
Matches 66; Conservative 63; Mismatches 99; Indels 32; Gaps 10;

QY 1 MEAKIDELINNDPVWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVWNLPKDQLAIV 60
Db 17 VNAALIEQLNSD-----VQLINQLGYIVSGGKRIRPMIAVLAARAVGEGNAHVIT 69

QY 61 SQIVELLHNSLLIDIEDNAPLRGQTTSHLIFGVPSITINTANYMYFRAMOLVSLQTTK 120
Db 70 AALIEFIHTATLLHDDVDVDESDMRRGKATANAAGNAASVLVGDFTYTRAFQMTSLGSL 129

QY 121 EPLYHNLTITFNEELINLHRGQGLDIY-WRDFLPEIPIPTQEMVYLNVMNKTGGLFELTLR 179
Db 130 K-----VLEWSEAVNVIAEGEVLQLMNVND--PDI--TEENYMRVIYKTAFLFEAAQ 180

QY 180 ---LMEALSPSSHGHSHLVPFNLGLIYQIRDDYLNLDKDFQMSSE---KGFAEDITEGK 233
Db 181 CSGIAGCTPEEKG--LDYGRYLGTAFLQIDLL--DYNADGEQLGNVGDLLNEGK 235

QY 234 LSPPIVHALNFTKGTQTEQ 253
Db 236 PTLPLLLHAMH-----HGTPEQ 251

Search completed: October 17, 2004, 08:54:19
Job time : 21.8543 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 08:53:05 ; Search time 31.4343 Seconds
(without alignments)
3445.655 Million cell updates/sec

Title: US-10-041-018-22

Perfect score: 1720

Sequence: 1 MEAKIDELINNDPVWSSQNE.....DTATNLHDELLYIDHLSL 335

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1360919 seqs, 323318874 residues

Total number of hits satisfying chosen parameters: 1360919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10F_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1720	100.0	335	14	US-10-369-493-22494
2	1720	100.0	335	15	US-10-450-941-6
3	1720	100.0	335	15	US-10-041-018-22
4	629.5	36.6	348	14	US-10-369-493-12405
5	623	36.2	428	14	US-10-369-493-4221
6	591.5	34.4	303	15	US-10-041-018-34
7	583	33.9	300	15	US-10-041-018-34
8	577	33.5	304	9	US-09-925-302-614
9	577	33.5	304	10	US-09-925-302-614
10	575.5	33.5	452	14	US-10-369-493-12982
11	545.5	31.7	338	15	US-10-041-018-24
12	545.5	31.7	338	15	US-10-041-018-77
13	538	31.3	376	15	US-10-001-192A-15
14	513	29.8	335	15	US-10-041-018-23

15	513	29.8	335	15	US-10-041-018-76	Sequence 76, Appl
16	490	28.5	319	14	US-10-369-493-12991	Sequence 12991, A
17	484	28.5	307	14	US-10-369-493-12663	Sequence 12663, A
18	384	22.3	304	14	US-10-369-493-12919	Sequence 12919, A
19	337	19.6	347	15	US-10-282-122A-47089	Sequence 47089, A
20	310.5	18.1	327	14	US-10-369-493-21492	Sequence 21492, A
21	274.5	16.0	364	15	US-10-282-122A-76572	Sequence 76572, A
22	246.5	14.3	322	15	US-10-282-122A-68191	Sequence 68191, A
23	243.5	14.2	323	15	US-10-282-122A-78209	Sequence 78209, A
24	242.5	14.1	307	15	US-10-282-122A-68693	Sequence 68693, A
25	239	13.9	325	14	US-10-369-493-1101	Sequence 1101, Ap
26	238.5	13.9	324	15	US-10-282-122A-66086	Sequence 66086, A
27	236	13.7	324	15	US-10-282-122A-65395	Sequence 65395, A
28	234	13.6	322	14	US-10-369-493-13946	Sequence 13946, A
29	233	13.5	322	9	US-09-815-242-12031	Sequence 12031, A
30	233	13.5	322	15	US-10-282-122A-66686	Sequence 66686, A
31	228.5	13.3	322	15	US-10-282-122A-69499	Sequence 69499, A
32	227	13.2	321	15	US-10-282-122A-51273	Sequence 51273, A
33	226.5	13.2	310	14	US-10-369-493-19696	Sequence 19696, A
34	226	13.1	323	9	US-09-815-242-13734	Sequence 13734, A
35	226	13.1	323	15	US-10-282-122A-75052	Sequence 75052, A
36	226	13.1	323	15	US-10-282-122A-75649	Sequence 75649, A
37	225	13.1	331	14	US-10-369-493-10088	Sequence 10088, A
38	223	13.0	323	9	US-09-815-242-10321	Sequence 10321, A
39	223	13.0	323	14	US-10-381-779-51	Sequence 51, Appl
40	223	13.0	323	14	US-10-369-493-23568	Sequence 23568, A
41	223	13.0	323	15	US-10-282-122A-56708	Sequence 56708, A
42	220.5	12.8	305	15	US-10-282-122A-59402	Sequence 59402, A
43	219.5	12.8	311	14	US-10-369-493-10961	Sequence 10961, A
44	219.5	12.8	330	15	US-10-282-122A-66971	Sequence 66971, A
45	215.5	12.5	309	15	US-10-282-122A-50552	Sequence 50552, A

ALIGNMENTS

RESULT 1

US-10-369-493-22494
; Sequence 22494, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22494
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22494

Query Match 100.0%; Score 1720; DB 14; Length 335;
Best Local Similarity 100.0%; Pred. No. 2.1e-140;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MEAKIDELINNDPVWSSQNE	SLSKPNYHLLKPKGNFRNLIVQINRNVNLPKQQLAIV	60
Db	1	MEAKIDELINNDPVWSSQNE	SLSKPNYHLLKPKGNFRNLIVQINRNVNLPKQQLAIV	60
Qy	61	SQIVELHNSLLIDIEDNAPLRGQTTS	SHLIFGVFSTINTANTMYFRAMQVLSQITTK	120
Db	61	SQIVELHNSLLIDIEDNAPLRGQTTS	SHLIFGVFSTINTANTMYFRAMQVLSQITTK	120
Qy	121	EPLVHNLITFNEELINLHRGQGLDI	WRDFLPEIITQEMVLMNWNKGGGLFRULRL	180

Db 121 EPLVNLITIFNEELINLHROGGLDIYWRDPLPEIPIQEMVYLNVMNKTGGLFRLTLRL 180
QY 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPIVH 240
Db 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPIVH 240
QY 241 ALNFTKTKGQTEQHNEIRILLRTSDKDIKLIQILEFDTNSLAYTKNFQINQVNMV 300
Db 241 ALNFTKTKGQTEQHNEIRILLRTSDKDIKLIQILEFDTNSLAYTKNFQINQVNMV 300
QY 301 NDENKYLPLDASHSDTATNLHDELLYIIDHSEL 335
Db 301 NDENKYLPLDASHSDTATNLHDELLYIIDHSEL 335

RESULT 2
US-10-450-941-6
; Sequence 6, Application US/10450941
; Publication No. US20040063182A1
; GENERAL INFORMATION:
; APPLICANT: TOYOTA JIDOSHA KABUSHIKI KAISHA
; TITLE OF INVENTION: A METHOD OF PRODUCING PRENYLALCOHOL
; FILE REFERENCE: PH-1444PCT
; CURRENT APPLICATION NUMBER: US/10/450,941
; CURRENT FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: JP2000-401701
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: JP2000-403067
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: JP2001-282978
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 6
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-450-941-6

Query Match 100.0%; Score 1720; DB 15; Length 335;
Best Local Similarity 100.0%; Pred. No. 2.1e-140;
Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEAKIDELINNDPVMSSQNESLISKYPYNNHLLKPGKNFRLNLIVQINRVNMLPKQOLAIV 60
Db 1 MEAKIDELINNDPVMSSQNESLISKYPYNNHLLKPGKNFRLNLIVQINRVNMLPKQOLAIV 60
QY 61 SQIVELLHNSLLIDIEDNAPLRGQTTSHLIFGVPSSTINTANYMYFRAMQVLSQITTK 120
Db 61 SQIVELLHNSLLIDIEDNAPLRGQTTSHLIFGVPSSTINTANYMYFRAMQVLSQITTK 120
QY 121 EPLYHNLITIFNEELINLHROGGLDIYWRDPLPEIPIQEMVYLNVMNKTGGLFRLTLRL 180
Db 121 EPLYHNLITIFNEELINLHROGGLDIYWRDPLPEIPIQEMVYLNVMNKTGGLFRLTLRL 180
QY 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPIVH 240
Db 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPIVH 240
QY 241 ALNFTKTKGQTEQHNEIRILLRTSDKDIKLIQILEFDTNSLAYTKNFQINQVNMV 300
Db 241 ALNFTKTKGQTEQHNEIRILLRTSDKDIKLIQILEFDTNSLAYTKNFQINQVNMV 300
QY 301 NDENKYLPLDASHSDTATNLHDELLYIIDHSEL 335
Db 301 NDENKYLPLDASHSDTATNLHDELLYIIDHSEL 335

RESULT 3
US-10-041-018-22
; Sequence 22, Application US/10041018
; Publication No. US2004007232A1

; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 22
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-22

Query Match 100.0%; Score 1720; DB 15; Length 335;
Best Local Similarity 100.0%; Pred. No. 2.1e-140;
Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEAKIDELINNDPVMSSQNESLISKYPYNNHLLKPGKNFRLNLIVQINRVNMLPKQOLAIV 60
Db 1 MEAKIDELINNDPVMSSQNESLISKYPYNNHLLKPGKNFRLNLIVQINRVNMLPKQOLAIV 60
QY 61 SQIVELLHNSLLIDIEDNAPLRGQTTSHLIFGVPSSTINTANYMYFRAMQVLSQITTK 120
Db 61 SQIVELLHNSLLIDIEDNAPLRGQTTSHLIFGVPSSTINTANYMYFRAMQVLSQITTK 120
QY 121 EPLYHNLITIFNEELINLHROGGLDIYWRDPLPEIPIQEMVYLNVMNKTGGLFRLTLRL 180
Db 121 EPLYHNLITIFNEELINLHROGGLDIYWRDPLPEIPIQEMVYLNVMNKTGGLFRLTLRL 180
QY 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPIVH 240
Db 181 MEALSPSSHGHSLVFPFNLGLIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPIVH 240
QY 241 ALNFTKTKGQTEQHNEIRILLRTSDKDIKLIQILEFDTNSLAYTKNFQINQVNMV 300
Db 241 ALNFTKTKGQTEQHNEIRILLRTSDKDIKLIQILEFDTNSLAYTKNFQINQVNMV 300
QY 301 NDENKYLPLDASHSDTATNLHDELLYIIDHSEL 335
Db 301 NDENKYLPLDASHSDTATNLHDELLYIIDHSEL 335

RESULT 4
US-10-369-493-12405
; Sequence 12405, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12405
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-369-493-12405

Query Match 36.6%; Score 629.5; DB 14; Length 348;
Best Local Similarity 43.2%; Pred. No. 5.2e-46;
Matches 127; Conservative 63; Mismatches 87; Indels 17; Gaps 6;

Qy	2	EAKIDELINDPVSSQNESLISKYPHNHILLKPKGNFRNLIVQINRVMNLKPKQLAIVS	61
Db	53	QSSLDGTYKXGKQKNSQNEEVIWGPYDYMHPQKDLRQMFNAFVNWLVKPSSESLAIT	112
Qy	62	QIVELLHNSLLIIDIEDNAPLRGQTTSHLLFGVPSSTINTANYFRAVOLVSQLTKE	121
Db	113	KVAMLHTASLLIDVEDNSLRLGPIVAHSYIGTAQTINSANYVFLALQEVOKU--KS	170
Qy	122	PLYHNLTIFNEELINLHRGQGLDIWYRDFLEPIITQBYMLNMVNKTKGGLFRLTRLM	181
Db	171	PA---AIDIVYOELLNLHRGQGMDFWRDTL--TCPSEDEYLEVMGNTKGGFLFRAVKLM	225
Qy	182	BALSPSSHGHGSLVPFINLGHIIYQIRDDYLNLDKQFMSSEKGFADITEGKLSPIVHA	241
Db	226	QA---ESSTGKDCVALVNVGLVFOCDYLNLSUTTYTQNKGLCEDLTEGKFSPLIHS	282
Qy	242	LNFTYKTKQTEQHNEILRIILLRTSDKOIKLKIQILEPDTNSLAYTKNFINQL	295
Db	283	IR-----SNPGNIGLINLRQTKDEEYKVALQYME-STGSKHTQDWQRQL	329

```

RESULT 5
US-10-369-493-4221
; Sequence 4221, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 4221
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-369-493-4221

```

Query Match.	36.2%;	Score 623;	DB 14;	Length 428;
Best Local Similarity	39.1%;	Pred. No. 2.5e-45;		
Matches	125;	Conservative	74;	Mismatches 99; Indels 22; Gaps 7
QY	15	WSSQNEGLISKPNHLLKCKNFRLNLIVGINRVMLPKDQLAIVSQIVBLLHNSLLI	74	
Db	121	WSSEKVLTPDYDLNGHGKDIRSQWKAFDALWDVPSSLEVIITKVISMLTASLLV	180	
QY	75	DDIEDNAPLRGQTTSHLIGVPSPTINTANYMYFRAMQLVSQLTTKPLVHNLITIFNEE	134	
Db	181	DDVEDNSVLRGPPVAHSIGIPQTINTSNVYFYALQELKL---KVP---KAVSIIFSE	235	
QY	135	LINLHRGQGLDIYWRDPLPEIIPQENMYLNMVNMKTGGLFRLTLRLMEALSPSHHGSL	194	
Db	236	LLNLHRGQGLDFWRDYL--TCPTEDDYLEMVSNKTKGGLFRLGKLMQAESRP---VDC	290	
QY	195	VFINLLGILYQREDDYLNLLKDFQWSEKGFPAEDITTEKGLSPFPIVHAINFTKVGQTEQH	254	
Db	291	VPLVNIIGLIQFQADDYHNLWNREYTKANGKCEDLTGKGFSPFIHSIR-----SNPSN	344	
QY	255	NEIRILLRLTSDKDIKLTQLILEFDPTNSLAYTKNFINOLVNMIKNDNENKYLPLDLASH	314	
Db	345	MLLNILKQKTGDEEVKRYAVAYME-STGSFEYTRKVIKVLV-----DRAQWTDIDDG	398	
QY	315	SDTATNLHDBELLYIIDLHSE	334	
Db	399	RKSGGITHKILDRIMLHOEE	418	

```

RESULT 6
US-10-041-018-34
; Sequence 34, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259980
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Mucor circinelloides f. lusitanicus
US-10-041-018-34

```

```

Query Match      34.4%; Score 591.5; DB 15; Length 303;
Best Local Similarity 41.1%; Pred. No. 8.3e-43;
Matches 113; Conservative 67; Mismatches 78; Indels 17; Gaps 5

QY 20 ESLSKPYNHLLKPKGNFRNLIVQNRVWNLPKQQLAIVSQIVELHNSLLIDIED 79
Db 14 EDILPEPTYLSQPGKIRAKLISAFDLHLHPKDVLCVINKTIQGLHNASLIMDDVQ 73
QY 80 NAPLRGQTTSHLIPGVSTINTANYVFRAMQVLSQTTKEPLYHNLITIFNEELINLH 139
Db 74 DDLRGVPVAHHYGVGVQTINTANYVIFALQVEMKLNIP-----SMMQVCTEELINLH 128
QY 140 RGQGLDIYWRDPLPEIIPQEMYLNMVYKTTGGIFLRTLRLMEALSPSSHGHSLVPLFN 199
Db 129 RGQGIELYWRDSL--TCPTEEYIDWYKTSGLLRVLAVRLMQAESD---IDYTPLVN 183
QY 200 LGITIYQRDDYLNLKQFMSSEKGAEDITEGKLSFPPIVHALNFTKTKGQEHNEILR 259
Db 184 IIGIHQVRQDDYMNLOQSTYNNKGFCEDLTEGKFSPIIHAIr-----KQPSNRQLLN 237
QY 260 ILLRTSDKDKLKLILQLEPDTNSLAYTKNFNQ 294
Db 238 IISQKPTSEIVKVALEVIR-KAGSFYEVFRELRQ 271

```

```

RESULT 7
US-10-041-018-46
; Sequence 46, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Human
US-10-041-018-46

```

Query Match	33.9%;	Score 583;	DB 15;	Length 300;
Best Local Similarity	40.5%;	Pred. No. 4.5e-42;		
Matches 128; Conservative	55;	Mismatches 103;	Indels 30;	Gaps 8;

; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 12982
 ; LENGTH: 452
 ; TYPE: PRT
 ; ORGANISM: Aspergillus nidulans
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(452)
 ; OTHER INFORMATION: unsure at all Kaa locations
 US-10-369-493-12982

Query Match 33.5%; Score 575.5; DB 14; Length 452;
 Best Local Similarity 40.3%; Pred. No. 3.5e-41;
 Matches 133; Conservative 47; Mismatches 99; Indels 51; Gaps 6;

QY 10 NNDPFWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVWMLPKDQLAIVSQIVELLHNS 69
 DB 105 DNQHQKKEEIIISAPVHYIISLTPGDKIRGKLISAFNEWFQIPDEQLEIIKRAVGLHV 164
 QY 70 SLL-----IDDEDNAPLRGQTTSHLIFGVPSTINTANYM 106
 DB 165 ASLLKGSFTLTHQITQFQIDNLCRRIDDIEDYKLRGFPVAHSIFIGPQTINSANYA 224
 QY 107 YFRAMQLVQITTEPELVNLIITFNEELINHRGQGLDIYWRDPLPFIITQENYLMV 166
 DB 225 YFOQNEVLKLRCE---KALRIFTTELLHRGQGMIDYWRDSL--TCPSEBEYDMV 278
 QY 167 MNKGGGLFRLRLMEALSPSSHHGSL-----VPEINLIGIYO 206
 DB 279 ANKGGGLFRLAIKLIQLESVDVEFVRPLEFLPFXRAEQYKLRRDCVPLVDLGIIFQ 338
 QY 207 IRDYLNLKDFQMSSEKGFADITEGKLSFPFIVHALNFTKKG-QTEGHNEILRIILRT 265
 DB 339 IRDDYQNLQSDQYAKNKGFGEDITEGKFSYPIVHSIRSSRTGVSQSSQLLSILRQKT 398
 QY 266 SDKDKIKLQILFEDTNSLAYTKNFINQL 295
 DB 399 EDEAVKXYTIQILR-KTGSFEFTKQLRDL 427

RESULT 11
 US-10-041-018-24
 ; Sequence 24, Application US/10041018
 ; Publication No. US20040072323A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Matsuda, Seiichi P.T.
 ; APPLICANT: Hart, Elizabeth A.
 ; TITLE OF INVENTION: Biterpene-Producing Unicellular Organism
 ; FILE REFERENCE: P02080US/10025547
 ; CURRENT APPLICATION NUMBER: US/10/041,018
 ; CURRENT FILING DATE: 2002-01-07
 ; PRIOR APPLICATION NUMBER: US 60/259880
 ; PRIOR FILING DATE: 2001-01-05
 ; NUMBER OF SEQ ID NOS: 413
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 24
 ; LENGTH: 338
 ; TYPE: PRT
 ; ORGANISM: Drosohila melanogaster
 US-10-041-018-24

Query Match 31.7%; Score 545.5; DB 15; Length 338;
 Best Local Similarity 38.7%; Pred. No. 9.3e-39;
 Matches 123; Conservative 55; Mismatches 113; Indels 27; Gaps 7;

QY 12 DPVWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVWMLPKDQLAIVSQIVELLHNS 71
 DB 13 DKSTQKEDEILLQPFYIQQIPGKQFRSELAFNHWLLIPGKLAIGDIVQMLHNS 72
 QY 72 LLIDIEDNAPLRGQTTSHLIFGVPSTINTANYMFRAMQLVQITTEPELVNLIITIF 131
 DB 73 LLIDIEDNAPLRGQTTSHLIFGVPSTINTANYMFRAMQLVQITTEPELVNLIITIF 131
 QY 132 NEELINHRGQGLDIYWRDPLPFIITQENYLMVWMLPKDQLAIVSQIVELLHNS 191
 DB 128 TEQLLEHRGQGMELIYWRDSF--TCPSESDYKLMVTRKVTGGLFMAIRLMQIFSSNKEDY 185
 QY 192 HSLVPPFINLGIYQIRDDYCNLSKEYTENKSPAEDITEGKLSFPFIVHALNFTKQGT 251
 DB 186 SKLTA---ILGLYFQIRDDYCNLSKEYTENKSPAEDITEGKLSFPFIVHALNFTKQGT 237
 QY 252 EQHNEILRIILRTSDKDKIKLQILFEDTNSLAYTKNFINQLVNMINKND----NENKY 307
 DB 238 -QDKQVLHILRQTHDIEVKKYCITLLE-KLGSFQYTRKVLKESLDAEARESEVARLGSNPY 295
 QY 308 LPDLASH-----SDTAT 319
 DB 296 MDRLINKLLSWKTSDSAS 313

RESULT 12
 US-10-041-018-77
 ; Sequence 77, Application US/10041018
 ; Publication No. US20040072323A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Matsuda, Seiichi P.T.
 ; APPLICANT: Hart, Elizabeth A.
 ; TITLE OF INVENTION: Biterpene-Producing Unicellular Organism
 ; FILE REFERENCE: P02080US/10025547
 ; CURRENT APPLICATION NUMBER: US/10/041,018
 ; CURRENT FILING DATE: 2002-01-07
 ; PRIOR APPLICATION NUMBER: US 60/259880
 ; PRIOR FILING DATE: 2001-01-05
 ; NUMBER OF SEQ ID NOS: 413
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 77
 ; LENGTH: 338
 ; TYPE: PRT
 ; ORGANISM: Drosohila melanogaster
 US-10-041-018-77

Query Match 31.7%; Score 545.5; DB 15; Length 338;
 Best Local Similarity 38.7%; Pred. No. 9.3e-39;
 Matches 123; Conservative 55; Mismatches 113; Indels 27; Gaps 7;

QY 12 DPVWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVWMLPKDQLAIVSQIVELLHNS 71
 DB 13 DKSTQKEDEILLQPFYIQQIPGKQFRSELAFNHWLLIPGKLAIGDIVQMLHNS 72
 QY 72 LLIDIEDNAPLRGQTTSHLIFGVPSTINTANYMFRAMQLVQITTEPELVNLIITIF 131
 DB 73 LLIDIEDNAPLRGQTTSHLIFGVPSTINTANYMFRAMQLVQITTEPELVNLIITIF 131
 QY 132 NEELINHRGQGLDIYWRDPLPFIITQENYLMVWMLPKDQLAIVSQIVELLHNS 191
 DB 128 TEQLLEHRGQGMELIYWRDSF--TCPSESDYKLMVTRKVTGGLFMAIRLMQIFSSNKEDY 185
 QY 192 HSLVPPFINLGIYQIRDDYCNLSKEYTENKSPAEDITEGKLSFPFIVHALNFTKQGT 251
 DB 186 SKLTA---ILGLYFQIRDDYCNLSKEYTENKSPAEDITEGKLSFPFIVHALNFTKQGT 237
 QY 252 EQHNEILRIILRTSDKDKIKLQILFEDTNSLAYTKNFINQLVNMINKND----NENKY 307
 DB 238 -QDKQVLHILRQTHDIEVKKYCITLLE-KLGSFQYTRKVLKESLDAEARESEVARLGSNPY 295

QY 308 LPDLASH-----SDTAT 319
Db 296 MDRLNKLKLSWKTSDSAS 313

RESULT 13

US-10-001-192A-15
; Sequence 15, Application US/10001192A
; Publication No. US20040091958A1

GENERAL INFORMATION:

; APPLICANT: Ooijen, Albert

; APPLICANT: Verdoes, Jan

; APPLICANT: Wery, Jan

; TITLE OF INVENTION: IMPROVED METHODS FOR TRANSFORMING

; TITLE OF INVENTION: PHAFFIA STRAINS, TRANSFORMED PHAFFIA STRAINS SO OBTAINED AND

; TITLE OF INVENTION: RECOMBINANT DNA IN SAID METHODS

; FILE REFERENCE: 24615-20104.01

; CURRENT APPLICATION NUMBER: US/10/001,192A

; CURRENT FILING DATE: 2002-06-11

; PRIOR APPLICATION NUMBER: EP 95203620.0

; PRIOR FILING DATE: 1995-12-22

; PRIOR APPLICATION NUMBER: EP96200943.7

; PRIOR FILING DATE: 1996-04-11

; NUMBER OF SEQ ID NOS: 57

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15

; LENGTH: 376

; TYPE: PRT

; ORGANISM: Phaffia rhodozyma

US-10-001-192A-15

Query Match

Best Local Similarity 31.1%; Score 538; DB 15; Length 376;

Matches 117; Conservative 79; Mismatches 110; Indels 70; Gaps 7;

QY 15 WSSQNESLSKPYNHILLKPGKFNRLNLIQVNRVNNLPKQDLAIYSQIVVELHNSSLI 74
Db 14 FTQDDIVLLEPVHYLGKPGKEIRSQLEAFNWLDDVKEDLEVIQNVVGLHTASLLM 73

QY 75 DIIDNAPLRGGTTHLIFGVSTINTANYMYFRAMQLV-----SDTAT 319
Db 74 DDVEDSVLRGSPVAHLIYGIPQTINTANYVFLAYQEIFKLRPTPIPMVIPPSASL 133

QY 115 -----SQTTEKPYHNLITFNEELNLRGGQLDIYWRD 150
Db 134 QSVSVSASSSSASSNNGTSTPNFSKDTY--LDKVTDEMLSLRGGQLELFWRD 191

QY 151 FLPEIIPQEMLYNMVNMKTTGGLFRLLRLMEALSPSHHSHLVPFINLLGIYQIRD 210
Db 192 SL--TCPSEHYKVLKGTGGLFRIAVRLMA--KSECDIDFVQLVNLISYFQIRD 246

QY 211 YNLKDFQMSSEKGFADITEGKLSPTPIVHALNFTKGTQTEQHNEILRIILLRTSKDI 270
Db 247 YNNLQSEYAHNKNFAEDITEGKFSPTIHSIHANPSS-----RLVINTLQKKSTPEI 300

QY 271 KLKLIQILEFDTNSLAYTKNFQINQVNMKND-----NENKYLPLDLASHSDTAT 319
Db 301 LHCVMYMTTSHSFYQVILNTLSGALRELGRLOGEFAEANSRMDLGDVDSEGTGK 360

QY 320 NLHDELLYIIDLSEL 335
Db 361 NV--KLEAILKKLADI 374

RESULT 14

US-10-041-018-23

; Sequence 23, Application US/10041018

; Publication No. US20040072323A1

GENERAL INFORMATION:

; APPLICANT: Matsuda, Seiichi P.T.

; APPLICANT: Hart, Elizabeth A.

; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism

; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018

; CURRENT FILING DATE: 2002-01-07

; PRIOR APPLICATION NUMBER: US 60/259880

; PRIOR FILING DATE: 2001-01-05

; NUMBER OF SEQ ID NOS: 413

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 23

; LENGTH: 335

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-10-041-018-23

Query Match 29.8%; Score 513; DB 15; Length 335;

Best Local Similarity 37.7%; Pred. No. 5.9e-36;

Matches 120; Conservative 56; Mismatches 112; Indels 30; Gaps 8;

QY 12 DPVWSSQNESLSKPYNHILLKPGKFNRLNLIQVNRVNNLPKQDLAIYSQIVVELHNS 71
Db 13 DKSTQKEQDEILLQPTTYIQIPGKQFRSELALAFNHWLLIPGEKLAQIGNIVQMLHNS 72

QY 72 LLIDDIEDNAPLRGGTTHLIFGVSTINTANYMYFRAMQLVSQLTTEKPYHNLITIF 131
Db 73 LL---IEDNSILRRGVPAHSIYGVASTINAAANYALFLALEKVQQDDHEF-----VRKY 124

QY 132 NEELINLRGGQLDIYWRDLPFELPIIPTQEMLYNMVNMKTTGGLFRLLRLMEALSPSHHG 191
Db 125 TEQLELHRGGQWEIYWRDSF--TCPSESDYKLMVTKTGGGLFRLMLAQLFSSNKEDY 182

QY 192 HSLVFPINLLGIYQIRDDYLNKDFQMSSEKGFADITEGKLSPTPIVHALNFTKGT 251
Db 183 SKLTA---ILGLYFQIRDDYCNLSKEVSNKSFADLTEGKFGFPIHNAVTKQ----- 234

QY 252 EOHNETILRLITSDKDIKLIQILEFDTNSLAYTKNFQINQVNMKND-----NENKY 307
Db 235 -QKQVLRKSSORTHDIEVKYCIITLLE-KLGSFQYTRKVLSDAEARSEVSTLGSN 292

QY 308 LPDLASH-----SDTAT 319
Db 293 MDRLNKLKLSWKTSDSAS 310

RESULT 15

US-10-041-018-76

; Sequence 76, Application US/10041018

; Publication No. US20040072323A1

; GENERAL INFORMATION:

; APPLICANT: Matsuda, Seiichi P.T.

; APPLICANT: Hart, Elizabeth A.

; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism

; FILE REFERENCE: P02080US1/10025547

; CURRENT APPLICATION NUMBER: US/10/041,018

; CURRENT FILING DATE: 2002-01-07

; PRIOR APPLICATION NUMBER: US 60/259880

; PRIOR FILING DATE: 2001-01-05

; NUMBER OF SEQ ID NOS: 413

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 76

; LENGTH: 335

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-10-041-018-76

Query Match 29.8%; Score 513; DB 15; Length 335;

Best Local Similarity 37.7%; Pred. No. 5.9e-36;

Matches 120; Conservative 56; Mismatches 112; Indels 30; Gaps 8;

QY 12 DPVWSSQNESLSKPYNHILLKPGKFNRLNLIQVNRVNNLPKQDLAIYSQIVVELHNS 71
Db 13 DKSTQKEQDEILLQPTTYIQIPGKQFRSELALAFNHWLLIPGEKLAQIGNIVQMLHNS 72

QY 72 LLIDDIEDNAPLRGGTTHLIFGVSTINTANYMYFRAMQLVSQLTTEKPYHNLITIF 131
Db 73 LL---IEDNSILRRGVPAHSIYGVASTINAAANYALFLALEKVQQDDHEF-----VRKY 124

Db 73 LL---IEDNSILRRGYPVAHSIYGVASTINAANYALFLALEKVQQLDHPE-----VRKY 124
Qy 132 NEELINLRGQGLDIYWRDFLPEIIPTOEWYLMVMNKTGGLFRLLRLMEALSPSSHG 191
Db 125 TEQLLELRGQGWELIWRDSF--TCPSESDYKLMVVRKTGGGLFLAIRLMQLFSSNKEDY 182
Qy 192 HSLVFPINLLGIYQIRDDYLNLDKDFQMSSEKGFADITEGKLSPPIVHALNFTTKGT 251
Db 183 SKLTA---ILGLYFQIRDDYCNLSKEVSNKSFADLTEGKFGFPVHAVTQK----- 234
Qy 252 EQHNEILRILLRTSDKDKIKLQILEPDTNSLAYTKNFQNLVNMKN-----NENKY 307
Db 235 -QDKQVLRKSQRTHDIEVKYCITILE-KLGSFOYTRKVLSELDAAERSEVSTLGSNPY 292
Qy 308 LPDLASH-----SDTAT 319
Db 293 MDRLINKLLSWKTSDSAS 310

Search completed: October 17, 2004, 09:18:43
Job time : 33.4343 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2004, 03:24:39 ; Search time 89,8123 Seconds
(without alignments)
1338.060 Million cell updates/sec

Title: US-10-041-018-22
Perfect score: 1720
Sequence: 1 MEAKIDELINNDPVWSSQNE.....DTATNLHDELIIIDHLSL 335

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1720	100.0	335	5	AAC22195
2	1720	100.0	335	5	ABG68086
3	1720	100.0	335	7	ADG49389
4	1720	100.0	335	8	ADL14811
5	1720	100.0	335	8	ADM98602
6	753	43.8	320	7	ADP75081
7	612.5	35.6	343	6	ADP96126
8	591.5	34.4	303	8	ADM98614
9	583	33.9	300	3	AAV68909
10	583	33.9	300	8	ADJ10383
11	583	33.9	300	8	ADM98626
12	577	33.5	304	3	AAV58276
13	574	33.4	300	2	AAV59753
14	574	33.4	300	2	AAV27005
15	570	33.1	300	2	AAV97565
16	545.5	31.7	338	4	ABB60947
17	545.5	31.7	338	8	ADM98657
18	545.5	31.7	338	8	ADM98604
19	538	31.3	376	2	AAW22498
20	513	29.8	335	8	ADM98656
21	513	29.8	335	8	ADM98603
22	437.5	25.4	320	4	AAU30743
23	337	19.6	347	6	ABU19165
24	274.5	16.0	364	6	ABU48648
25	246.5	14.3	322	6	ABU40267

26	243.5	14.2	323	6	ABU50285
27	242.5	14.1	307	6	ABU40769
28	242.5	14.1	307	7	ADF05623
29	239	13.9	325	7	ADB74153
30	238.5	13.9	324	6	ABU38162
31	236.5	13.8	330	2	AAW12391
32	236	13.7	324	6	ABP78926
33	236	13.7	324	6	ABU37471
34	233	13.5	322	4	AAU36438
35	233	13.5	322	6	ABU38762
36	233	13.5	341	7	ABO72702
37	231.5	13.5	330	2	AAW12389
38	230.5	13.4	324	6	ABM68417
39	229.5	13.3	330	2	AAW12386
40	229.5	13.3	330	2	AAW12390
41	228.5	13.3	322	6	ABU41575
42	228.5	13.3	330	2	AAV79969
43	228.5	13.3	330	2	AAW52283
44	228.5	13.3	330	2	AAW52282
45	228.5	13.3	330	2	AAW62528

ALIGNMENTS

RESULT 1
AAO22195
ID AAO22195 standard; protein, 335 AA.
XX AC AAO22195;
XX
XX
DT 03-OCT-2002 (first entry)
XX
DE Protein relating to the production of prenyl alcohol SEQ ID No 6.
XX
KW Prenyl alcohol; mutated cell; squalene synthase gene; geometric isomer;
KW industrial synthesis; isoprenoid-terpenoid compound.
XX
XX Saccharomyces cerevisiae.
XX
XX WO200253747-A1.
XX
XX 11-JUL-2002.
XX
XX 20-DEC-2001; 2001WO-JF011215.
XX
XX 28-DEC-2000; 2000JP-00401701.
XX
XX 28-DEC-2000; 2000JP-00403067.
XX
XX 18-SEP-2001; 2001JP-00282978.
XX
XX (TOYT) TOYOTA JIDOSHA KK.
XX
XX Ohto C, Obata S;
XX
XX WPI; 2002-548086/58.
XX
XX N-PSDB; AAL40786.
XX
XX Production of prenyl alcohols by culturing translationally-active mutated
XX cells with reduced squalene synthase gene to express less transcriptional
XX product.
XX
XX Claim 16; Page 146-148; 262pp; Japanese.
XX

The invention relates to a method for producing prenyl alcohol comprising culturing mutated cells, having been mutated so as to reduce the amount of transcriptional product of squalene synthase gene transcriptional activity, and then collecting prenyl alcohol from the culture medium. The method is for the production of prenyl alcohols, which is for use in industrial synthesis of isoprenoid-terpenoid compounds particularly physiologically-active prenyl alcohol geometric isomers. This sequence represents a protein relating to the method for producing prenyl alcohol comprising culturing mutated cells

SQ Sequence 335 AA;
 Query Match 100.0%; Score 1720; DB 5; Length 335;
 Best Local Similarity 100.0%; Pred. No. 4.6e-154;
 Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAKIDELINNDPWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVWMLPKDQLAIV 60
 DB 1 MEAKIDELINNDPWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVWMLPKDQLAIV 60

QY 61 SQIVELLHNSLLIIDDIEDNAPLRGGTTSLLIFGVPTSTINTANTMYFRAMOLVSQLT 120
 DB 61 SQIVELLHNSLLIIDDIEDNAPLRGGTTSLLIFGVPTSTINTANTMYFRAMOLVSQLT 120

QY 121 EPLYHNLITIFNEELINLHGGGLDIYWRDFLPEIIPQEMYLNMVNMKTGGLFRLTLRL 180
 DB 121 EPLYHNLITIFNEELINLHGGGLDIYWRDFLPEIIPQEMYLNMVNMKTGGLFRLTLRL 180

QY 181 MEALSPSSHGHGSLVPPFNILGIIYQIRDDYLNLDKDFQMSSEKGAEDITEGKLSPIVH 240
 DB 181 MEALSPSSHGHGSLVPPFNILGIIYQIRDDYLNLDKDFQMSSEKGAEDITEGKLSPIVH 240

QY 241 ALNFTKTKGTQEQHNEILLRLTSDKDIKLLIQLLEFDTNSLAYTKNFNLQVNMK 300
 DB 241 ALNFTKTKGTQEQHNEILLRLTSDKDIKLLIQLLEFDTNSLAYTKNFNLQVNMK 300

QY 301 NDENKYLPLASHSDTATNLHDELLIYIDHLSL 335
 DB 301 NDENKYLPLASHSDTATNLHDELLIYIDHLSL 335

RESULT 2
 ABG68086
 ID ABG68086 standard; protein; 335 AA.
 AC ABG68086;
 DE
 DT 24-SEP-2002 (first entry)
 XX
 XX S. cerevisiae prenyl diphosphate synthase #2.
 XX Prenyl alcohol; prenyl diphosphate synthase; geranylgeraniol;
 KW hydroxymethylglutaryl-CoA reductase; farnesylgeraniol; enzyme;
 KW isopentenyl diphosphate delta-isomerase; mevalonate kinase;
 KW mevalonate CoA acetyltransferase; isoprenoid-terpenoid compound.
 XX
 OS Saccharomyces cerevisiae.
 XX
 XX WO200253746-A1.
 XX
 XX 11-JUL-2002.
 XX
 XX 20-DEC-2001; 2001WO-JP011214.
 XX
 XX 28-DEC-2000; 2000JP-00403067.
 XX
 XX (TOYT) TOYOTA JIDOSHA KK.
 XX
 XX Ohto C, Obata S, Muramatsu M, Nishi K, Totsuka K;
 XX WPI; 2002-537944/57.
 XX
 XX Production of prenyl alcohols by culturing a transformant transferred
 PT with e.g. prenyl diphosphate synthase gene, for use in industrial
 PT synthesis of e.g. physiologically-active isoprenoid-terpenoid compounds.
 XX
 XX Disclosure; Page 185-187; 335pp; Japanese.
 XX
 XX The invention relates to a process for producing a prenyl alcohol
 CC comprising: (A) construction of a recombinant by transferring an
 CC expression recombinant DNA or a DNA for genome integration into a host
 CC which contains prenyl diphosphate synthase gene or its variant; and (B)
 CC collecting product from the culture medium. Also described is: (1) a

CC method for producing prenyl alcohol in which the expression recombinant
 CC DNA contains: (a) a hydroxymethylglutaryl-CoA reductase gene or its
 CC variant; or (b) an isopentenyl diphosphate delta-isomerase gene; (2) a
 CC process for producing geranylgeraniol, comprising: (a) constructing a
 CC recombinant by transferring an expression recombinant DNA or a DNA for
 CC genome integration into a host which contains hydroxymethylglutaryl-CoA
 CC reductase gene or its variant; and (b) isolating the product; (3) a
 CC process for producing farnesylgeraniol in which the expression
 CC recombinant DNA also contains an isopentenyl diphosphate delta-isomerase
 CC gene, mevalonate CoA acetyltransferase gene, hydroxymethylglutaryl-CoA
 CC synthase gene, mevalonate kinase gene, or mevalonate diphosphate
 CC decarboxylase gene. The methods are used for the production of prenyl
 CC alcohols, particularly for use in industrial synthesis of isoprenoid-
 CC terpenoid compounds e.g. physiologically-active prenyl alcohols including
 CC geometric isomers. ABG68084-ABG68095 represent prenyl diphosphate
 CC synthase proteins used in the methods of the invention
 XX
 XX Sequence 335 AA;
 SQ
 Query Match 100.0%; Score 1720; DB 5; Length 335;
 Best Local Similarity 100.0%; Pred. No. 4.6e-154;
 Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAKIDELINNDPWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVWMLPKDQLAIV 60
 DB 1 MEAKIDELINNDPWSSQNESLISKPNHLLKPGKFNRLNLIQVNRVWMLPKDQLAIV 60

QY 61 SQIVELLHNSLLIIDDIEDNAPLRGGTTSLLIFGVPTSTINTANTMYFRAMOLVSQLT 120
 DB 61 SQIVELLHNSLLIIDDIEDNAPLRGGTTSLLIFGVPTSTINTANTMYFRAMOLVSQLT 120

QY 121 EPLYHNLITIFNEELINLHGGGLDIYWRDFLPEIIPQEMYLNMVNMKTGGLFRLTLRL 180
 DB 121 EPLYHNLITIFNEELINLHGGGLDIYWRDFLPEIIPQEMYLNMVNMKTGGLFRLTLRL 180

QY 181 MEALSPSSHGHGSLVPPFNILGIIYQIRDDYLNLDKDFQMSSEKGAEDITEGKLSPIVH 240
 DB 181 MEALSPSSHGHGSLVPPFNILGIIYQIRDDYLNLDKDFQMSSEKGAEDITEGKLSPIVH 240

QY 241 ALNFTKTKGTQEQHNEILLRLTSDKDIKLLIQLLEFDTNSLAYTKNFNLQVNMK 300
 DB 241 ALNFTKTKGTQEQHNEILLRLTSDKDIKLLIQLLEFDTNSLAYTKNFNLQVNMK 300

QY 301 NDENKYLPLASHSDTATNLHDELLIYIDHLSL 335
 DB 301 NDENKYLPLASHSDTATNLHDELLIYIDHLSL 335

RESULT 3
 ADC49389
 ID ADC49389 standard; protein; 335 AA.
 XX
 XX ADC49389;
 AC
 XX
 XX 18-DEC-2003 (first entry)
 DT
 XX
 XX Yeast BTS1 protein.
 XX
 XX Yeast; microbiological; prenyl alcohol; geranylgeraniol; farnesol;
 KW nerolidol; squalene synthetase inhibitor; BTS1.
 XX
 XX Saccharomyces cerevisiae.
 XX
 XX JP2002300896-A.
 PN
 XX
 XX 15-OCT-2002.
 PD
 XX
 XX 18-JAN-2002; 2002JP-00010528.
 XX
 XX 30-JAN-2001; 2001JP-00021547.
 XX
 XX (TOYT) TOYOTA JIDOSHA KK.
 XX
 XX

DR WPI; 2003-485780/46.
 DR N-PSDB; ADC49385.
 XX
 PT A process for preparation of prenyl alcohols by culture of their
 PT productive microorganisms in a medium with added squalene synthetase
 PT inhibitor.
 XX
 PS
 PS Disclosure; SEQ ID NO 12; 37pp; Japanese.
 XX
 CC The invention discloses a microbiological preparation of prenyl alcohols,
 CC particularly geranylgeraniol, farnesol and/or nerolidol, by culture of
 CC their productive microorganism spp. of Saccharomyces, Saccharomycopsis,
 CC Saccharomycodes, Schizosaccharomyces, Wickerhamia, Debaryomyces,
 CC Hanseniaspora, Pichia, Candida, Zygosaccharomyces, Ogataea, Kuraihia,
 CC Komagataella, Yarrowia, Kluyveromyces, Trichosporon, Cryptococcus,
 CC Torulaspora, Bacillus, Staphylococcus, Micrococcus, Exiguobacterium,
 CC Gibberella or Mucor, or Ambrosiozyma, Cytotholobasidium, Metschnikowia,
 CC Trichosporiella, Xanthophyllomyces, Bullera, Fellomyces, Filobasidium,
 CC Holtermannia, Phaffia, Rhodotorula, Sporidiobolus, Sporobolomyces,
 CC Williopsis, Zygosascus, Leucosporidium, Myxozyma, Klöckera,
 CC Brevibacterium, Alcaligenes or Haloferax in a medium with added squalene
 CC synthetase inhibitor to extra- and intra-cellular cumulate prenyl
 CC alcohols, followed by their collection. The sequence presented is the
 XX yeast HWG1 cDNA.
 XX
 SQ Sequence 335 AA;
 Query Match 100.0%; Score 1720; DB 7; Length 335;
 Best Local Similarity 100.0%; Pred. No. 4.6e-154;
 Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MEAKIDELINNDPVWSSQNESLISKPNHLLKPGKNFRNLIVQINRVNMLPKDQLATV 60
 DB 1 MEAKIDELINNDPVWSSQNESLISKPNHLLKPGKNFRNLIVQINRVNMLPKDQLATV 60
 QY 61 SQIVELLHNSLLDIEDNAPLRGQTTSHLIFGVPTINTANTMYFRAMQLVSQLTTK 120
 DB 61 SQIVELLHNSLLDIEDNAPLRGQTTSHLIFGVPTINTANTMYFRAMQLVSQLTTK 120
 QY 121 EPLVHNLITIFNEELINLHRRGGGLDYWRDFELPEIIPQEWYLNVMVKTGGLFRLTLRL 180
 DB 121 EPLVHNLITIFNEELINLHRRGGGLDYWRDFELPEIIPQEWYLNVMVKTGGLFRLTLRL 180
 QY 181 MEALSPSSHHGSLVFFINLLGIYQIRDDYLNLDKDFQMSSEKGFPAEDITRGKLSFPIVH 240
 DB 181 MEALSPSSHHGSLVFFINLLGIYQIRDDYLNLDKDFQMSSEKGFPAEDITRGKLSFPIVH 240
 QY 241 ALNFTKTKGQTEQHNELIURLLLRTSDKDIKLLIQILEFDTNSLAYTKNFINQLVNMV 300
 DB 241 ALNFTKTKGQTEQHNELIURLLLRTSDKDIKLLIQILEFDTNSLAYTKNFINQLVNMV 300
 QY 301 NDNENKYLPLDASHSDTATNLHDELLYIIDHLSEL 335
 DB 301 NDNENKYLPLDASHSDTATNLHDELLYIIDHLSEL 335
 RESULT 4
 ID ADL14811
 XX ADL14811 standard; protein; 335 AA.
 XX
 XX ADL14811;
 DT 17-JUN-2004 (first entry)
 XX
 DE Yeast geranylgeranyl diphosphate synthase.
 XX
 KW Yeast; enzyme; Isopentenyl diphosphate biosynthetic pathway; IPP;
 KW prenyl alcohol; squalene synthase.
 XX
 OS Saccharomyces cerevisiae.
 XX
 XX US2004063182-A1.
 XX

PD 01-APR-2004.
 XX
 PF 18-JUN-2003; 2003US-00450941.
 XX
 PR 28-DEC-2000; 2000JP-00401701.
 PR 28-DEC-2000; 2000JP-00406067.
 PR 18-SEP-2001; 2001JP-00282978.
 PR 20-DEC-2001; 2001WO-JF011215.
 XX
 XX (OHTO/) OHTO C.
 PA (OBAT/) OBATA S.
 PA
 XX Ohto C, Ohta S;
 XX
 XX WPI; 2004-294401/27.
 DR N-PSDB; ADL14810.
 DR
 PT Producing prenyl alcohols by culturing a mutant cell in which the amount
 PT of squalene synthase gene transcript having translational activity is
 PT reduced.
 XX
 PS Claim 16; SEQ ID NO 6; 139pp; English.
 XX
 CC The invention relates to producing a prenyl alcohol comprising culturing
 CC a mutant cell that has been mutated so that an amount of squalene
 CC synthase gene transcript having translational activity can be reduced,
 CC and recovering the prenyl alcohol from the resultant culture.
 CC Alternatively, producing a prenyl alcohol also comprises culturing a
 CC mutant cell in which the transcription promoter region of its squalene
 CC synthase gene has been replaced with a transcription repression-type
 CC promoter under transcription repression conditions, reducing an amount of
 CC squalene synthase gene transcript having translational activity, and
 CC recovering the prenyl alcohol from the resultant culture. Also included
 CC are a method of producing a prenyl alcohol (comprising preparing a
 CC recombinant by introducing a recombinant DNA for expression or a DNA for
 CC genomic integration each comprising an IPP (isoprenyl diphosphate)
 CC biosynthetic pathway-related enzyme gene into any of the mutant cells
 CC cited above), a mutant cell that has been mutated so that an amount of
 CC squalene synthase gene transcript having translational activity can be
 CC reduced) and a mutant cell in which the transcription promoter region of
 CC its squalene synthase gene has been replaced with a transcription
 CC repression-type promoter so that an amount of transcript from its
 CC squalene synthase gene having translational activity can be reduced. The
 CC transcription repression-type promoter in producing a prenyl alcohol is
 CC GAL1 promoter. The IPP biosynthetic pathway-related enzyme gene is any
 CC one selected from farnesyl diphosphate synthase gene, geranylgeranyl
 CC diphosphate synthase gene, hydroxymethylglutaryl-CoA reductase gene,
 CC isopentenyl diphosphate A-isomerase gene, nevalonate kinase gene, acetyl-
 CC CoA acetyltransferase gene, hydroxymethylglutaryl-CoA synthase gene,
 CC phosphomevalonate kinase gene and diphosphomevalonate decarboxylase gene,
 CC or a fusion gene composed of a gene selected from the above genes, or a
 CC gene obtained by introducing an addition, substitution or insertion
 CC mutation into any one of the above genes so that the polypeptide encoded
 CC by the resultant gene contains an endoplasmic reticulum signal. The
 CC methods and compositions of the present invention are useful for
 CC producing prenyl alcohols by reducing the amount of squalene synthase
 CC gene transcript having translational activity. The present sequence
 CC represents an IPP (isoprenyl diphosphate) biosynthetic pathway-related
 CC enzyme.
 XX
 XX Sequence 335 AA;
 Query Match 100.0%; Score 1720; DB 8; Length 335;
 Best Local Similarity 100.0%; Pred. No. 4.6e-154;
 Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MEAKIDELINNDPVWSSQNESLISKPNHLLKPGKNFRNLIVQINRVNMLPKDQLAIV 60
 DB 1 MEAKIDELINNDPVWSSQNESLISKPNHLLKPGKNFRNLIVQINRVNMLPKDQLAIV 60
 QY 61 SQIVELLHNSLLDIEDNAPLRGQTTSHLIFGVPTINTANTMYFRAMQLVSQLTTK 120
 DB 61 SQIVELLHNSLLDIEDNAPLRGQTTSHLIFGVPTINTANTMYFRAMQLVSQLTTK 120

121 EPLVHNLITIFNEELINLHRCGLDIYWRDLPFLIIPTQEMVLNVMNKTGGLFRLTLRL 180
 121 EPLVHNLITIFNEELINLHRCGLDIYWRDLPFLIIPTQEMVLNVMNKTGGLFRLTLRL 180
 181 MEALSPSSHGHSLVFFNLGIIYQIRDDYLNKDFQMSSEKGFPAEDITEGKLSFPIVH 240
 181 MEALSPSSHGHSLVFFNLGIIYQIRDDYLNKDFQMSSEKGFPAEDITEGKLSFPIVH 240
 241 ALNFTTKGTQTEHNEILRLRTSDKDIKLIQILEFDNLSLAYTKNFNQLVNMWK 300
 241 ALNFTTKGTQTEHNEILRLRTSDKDIKLIQILEFDNLSLAYTKNFNQLVNMWK 300
 301 NDNENKYLPLDASHSDTATNLHDELLYIIDHLSSEL 335
 301 NDNENKYLPLDASHSDTATNLHDELLYIIDHLSSEL 335

RESULT 5
 ADM98602
 ID ADM98602 standard; protein; 335 AA.
 AC ADM98602;
 XX
 DT 01-JUL-2004 (first entry)
 XX Geranylgeranyl pyrophosphate synthase polypeptide #1.
 DE Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;
 KW diterpene synthase; defence toxin; volatile defensive signal;
 KW pollinator attractant; photoprotectant; enzyme.
 XX Saccharomyces cerevisiae.
 OS
 XX
 PN US2004072323-A1.
 XX
 PD 15-APR-2004.
 XX
 PF 07-JAN-2002; 2002US-00041018.
 XX
 PR 05-JAN-2001; 2001US-0259860P.
 XX
 PA (MATS/) MATSUDA S P T.
 PA (HART/) HART E A.
 XX
 PI Matsuda SPT, Hart EA;
 XX
 DR WPI; 2004-373921/35.
 XX
 XX New unicellular organisms comprising exogenous nucleic acids encoding a
 PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for
 PT producing diterpenes and diterpene precursors.

Claim 30; SEQ ID NO 22; 38pp; English.
 XX
 XX The invention relates to a unicellular organism for producing a diterpene
 CC or diterpene precursor comprising an exogenous nucleic acid sequence
 CC encoding a geranylgeranyl pyrophosphate synthase under the control of a
 CC promoter operable in the organism, and an exogenous nucleic acid sequence
 CC encoding a diterpene synthase under the control of a promoter operable in
 CC the organism. The invention also relates to methods of producing a
 CC diterpene or diterpene precursor and a method of isolating a diterpene
 CC synthase comprising growing several cells in the presence of a
 CC polyaromatic resin to make a cell/resin mixture, where at least one of
 CC the cells further comprises at least one isolated and purified nucleic
 CC acid sequence of a yeast expression library, and the expression of the
 CC nucleic acid sequence is regulated by an inducible promoter under
 CC conditions where the expression is induced, filtering the cell/resin
 CC mixture, extracting the cell/resin mixture with alcohol to produce an
 CC organic eluent and analysing the organic eluent by a screening method
 CC including chromatography and/or spectroscopy, to identify the nucleic
 CC acid sequence encoding the diterpene synthase. The unicellular
 CC microorganism is useful as a diterpene or diterpene precursor producing

CC system. Diterpenes, in plants, serve as defence toxins, volatile
 CC defensive signals, pollinator attractants and photoprotectants. This
 CC sequence represents a geranylgeranyl pyrophosphate synthase polypeptide
 CC of the invention. Note: The sequence data for this patent did not form
 CC part of the printed specification but was obtained in electronic format
 CC from USPTO at seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 335 AA;
 Query Match 100.0%; Score 1720; DB 8; Length 335;
 Best Local Similarity 100.0%; Pred. No. 4.6e-154;
 Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MEAKIDELINNDPVWSSQNESLISKPYNHILKPGKNFRLNLIQVNRVNNLPKQQLAIV 60
 DB 1 MEAKIDELINNDPVWSSQNESLISKPYNHILKPGKNFRLNLIQVNRVNNLPKQQLAIV 60
 QY 61 SQIVELLHNSLLIIDDIEDNAPLRGQTTSHLIFGVPTINTANYMYFAMQVLSQLTTK 120
 DB 61 SQIVELLHNSLLIIDDIEDNAPLRGQTTSHLIFGVPTINTANYMYFAMQVLSQLTTK 120
 QY 121 EPLVHNLITIFNEELINLHRCGLDIYWRDLPFLIIPTQEMVLNVMNKTGGLFRLTLRL 180
 DB 121 EPLVHNLITIFNEELINLHRCGLDIYWRDLPFLIIPTQEMVLNVMNKTGGLFRLTLRL 180
 QY 181 MEALSPSSHGHSLVFFNLGIIYQIRDDYLNKDFQMSSEKGFPAEDITEGKLSFPIVH 240
 DB 181 MEALSPSSHGHSLVFFNLGIIYQIRDDYLNKDFQMSSEKGFPAEDITEGKLSFPIVH 240
 QY 241 ALNFTTKGTQTEHNEILRLRTSDKDIKLIQILEFDNLSLAYTKNFNQLVNMWK 300
 DB 241 ALNFTTKGTQTEHNEILRLRTSDKDIKLIQILEFDNLSLAYTKNFNQLVNMWK 300
 QY 301 NDNENKYLPLDASHSDTATNLHDELLYIIDHLSSEL 335
 DB 301 NDNENKYLPLDASHSDTATNLHDELLYIIDHLSSEL 335

RESULT 6
 ADF75081
 ID ADF75081 standard; protein; 320 AA.
 XX
 AC ADF75081;
 XX
 DT 26-FEB-2004 (first entry)
 XX
 DE A_gossypii protein essential for fungal growth & development SegID 22.
 XX
 KW fungal growth; fungicide; screening assay; fungicidal; crop.
 XX
 OS Bremothecium gossypii.
 XX
 PN WO2003072706-A2.
 XX
 PD 04-SEP-2003.
 XX
 PF 11-FEB-2003; 2003WO-US0004154.
 XX
 PR 22-FEB-2002; 2002US-0359557P.
 XX
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 PI Gaffney TD, Flavler A, Dietrich F, Philippsen P, Kirksey CMM;
 PI Gates K, Gardner J;
 DR WPI; 2003-721756/68.
 DR N-PSDB; ADF75080.
 XX
 XX Identifying a fungicidal compound comprising a polypeptide with
 PT the compound to be tested with the ability to bind, or to inhibit the
 PT activity of the polypeptide under conditions conducive to binding or
 PT inhibition.
 XX

Claim 13; SEQ ID NO 22; 127pp; English.

This invention relates to novel nucleic acid molecules, isolated from *Ashbya gossypii*, which encode proteins essential for fungal growth and development. Specifically, it refers to methods for identifying these proteins as fungicide targets by using gene disruption with a geneticin resistance gene. In particular, the nucleic acids encoding the proteins required for fungal growth provide the basis of screening assays designed to easily and rapidly identify inhibitory compounds that exhibit fungicidal activity. Accordingly, the present invention describes such methods to identify protein inhibitors that can be used as fungicides to suppress the growth of undesirable fungi in agronomically important crops such as maize, wheat, barley, cotton and sugar beet. This polypeptide sequence is a protein sequence essential for fungal viability, used in an exemplification of the invention.

Sequence 320 AA;

Query Match 43.8%; Score 753; DB 7; Length 320;

Best Local Similarity 48.0%; Pred. No. 1.6e-62; Matches 155; Conservative 61; Mismatches 87; Indels 20; Gaps 7;

14 WSSQESLISGKPNHLLKPKFRLNLIVQINRVMNLPKDLAIVSQIVELLHNSLL 73

13 IWTASQELLQPNHLVTPQKPNRNTLIRVNGFYGLSERQVAATVELVEMLVASLL 72

74 IDIEDNAPLRGQTTSHLIFGVSTINTANVMYFRAMOLVSQLTKEPL--YHNLTIF 131

73 IDIEDNSAWRGVAAHVVGSPWNTANTANMVFVSNLSLGQLAAQRPAGPLQDLKVF 132

132 NEELINHRGGLDIYWRDFLPEIIPQEMVYLMNMTGGLFRLTLRLMEALSPS-SHH 190

133 NEEMNHLRGGLDIYWRDTP--TVPSHDIYLRVMHKTGGLFRLTLRLMEALRGPDGP 190

191 GHSVLPFNILGIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPIVHALNFTKYGQ 250

191 GSILVPLSNLGLVYQVRDDYLNDSRMSENKGFADITEGKFSYPIIHGLQYARVH-D 249

251 TEQNEILRIILLRTSKDKIKLKIQLIEFTNSLAYTKNFINOLVMNKNENKYLDP 310

250 PAGYDFLVSLRQRTDITTKRVVRVYLVADYSGSLAYTKQRIELATLIK---TKYIP- 304

311 LASHSDATNLHDELYIIDLHS 333

305 -ASGT-----ELCNVIDSLT 318

RESULT 7

ABP96126

ID ABP96126 standard; protein; 343 AA.

XX

AC

ABP96126;

08-MAY-2003 (first entry)

XX

DE

Phoma betae geranylgeranyl diphosphate synthase SEQ ID NO:7.

XX

KW

Phoma betae; aphidicolin biosynthase cluster; aphidicolin; GDP synthase;

aphidicolan-1,6 beta-ol synthase; geranylgeranyl diphosphate synthase;

cytochrome P450 hydroxylase; P450-1; P450-2; transporter protein; TP;

DNA polymerase alpha-specific inhibitor; replication; enzyme.

XX

OS

Phoma betae.

XX

XX

WC2003000902-A1.

XX

XX

03-JAN-2003.

PD

XX

XX

02-MAY-2002; 2002WO-JP004381.

FF

XX

21-JUN-2001; 2001JP-00188465.

PR

XX

(NISC-) JAPAN SCI & TECHNOLOGY CORP.

PA

XX

PI

XX

DR

XX

DR

XX

PT

XX

PT

XX

PT

XX

PS

XX

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

Oikawa H, Sassa T, Toyomasu T;

WPI; 2003-175292/17.

N-PSDB; ABZ77169.

New gene encoding an aphidicolan-1,6beta-ol synthase protein for producing transformants to produce the DNA polymerase alpha-specific inhibitor aphidicolin, applicable in e.g. drugs for controlling replication of eukaryotes.

Claim 4; Page 20-21; 74pp; Japanese.

The present invention describes genes required in the production of aphidicolin, which is a DNA polymerase alpha-specific inhibitor which plays an important role in the replication of eukaryotes. Genes involved in an aphidicolin biosynthase cluster are: aphidicolan-1,6 beta-ol synthase; geranylgeranyl diphosphate (GDP) synthase; cytochrome P450 hydroxylase (P450-1); transporter protein (TP) and cytochrome P450 hydroxylase (P450-2). The genes are used for constructing transformants to produce a DNA polymerase alpha-specific inhibitor aphidicolin by fermentation, applicable in reagents and drugs for cell cycle problems during co-culturing or for controlling replication of eukaryotes. With the transformants, aphidicolin can be produced in large quantities, and with high yield. The present sequence represents GDP synthase isolated from *Phoma betae*, which is used in the exemplification of the present invention

Sequence 343 AA;

Query Match 35.6%; Score 612.5; DB 6; Length 343;

Best Local Similarity 43.0%; Pred. No. 3.6e-49;

Matches 123; Conservative 60; Mismatches 86; Indels 17; Gaps 6;

15 WSSQESLISGKPNHLLKPKFRLNLIVQINRVMNLPKDLAIVSQIVELLHNSLLI 74

48 WTSKVKVTVGYDIAAKFKGKVRTLLACFDEWLVQPPLESLEVIQVVMHTASLLI 107

75 DDIEDNAPLRGQTTSHLIFGVSTINTANVMYFRAMOLVSQLTKEPLVHNLTIFNEE 134

108 DDIDQNSLGRGKPVQAQNIPTGALTINSANVYFLALEKLSL--KNP--NITDIFTEE 162

135 LINLRGGLDIYWRDFLPEIIPQEMVYLMNMTGGLFRLTLRLMEALSPSSHGSL 194

163 LLRLHRQAMDLWRDYL--TCPTTEEFYFVANKTGLFRLMYRMKA---ESSMFDL 217

195 VPFINLGIYQIRDDYLNKDFQMSSEKGFADITEGKLSFPIVHALNFTKKGQEQH 254

218 LPVVLLGVIFQVDDYKNLCSREYKGLKGFEDLTEGKFSFVHSIR-----SNPED 271

255 NEILRIILLRTSKDKIKLKIQLIEFTNSLAYTKNFINOLVMNIMK 300

272 LQLHLVLOQSSNEHVKLYAIEIME-STGSLETKHVVENIVSQIQ 316

RESULT 8

ADM98614

ID ADM98614 standard; protein; 303 AA.

XX

AC

ADM98614;

01-JUL-2004 (first entry)

XX

DE

XX

KW

Geranylgeranyl pyrophosphate synthase polypeptide #13.

XX

KW

Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;

diterpene synthase; defence toxin; volatile defensive signal;

pollinator attractant; photoprotectant; enzyme.

XX

OS

Mucor circinelloides f. lusitanicus.

XX

XX

US2004072323-A1.

PD	15-APR-2004.	AA68909 standard; protein; 300 AA.	ID
XX			XX
PF	07-JAN-2002; 2002US-00041018.	AA68909;	AC
XX			XX
PR	05-JAN-2001; 2001US-0259880P.	16-MAY-2000 (first entry)	DT
XX			XX
PA	(MATS/) MATSUDA S P T.	A geranylgeranyl pyrophosphate synthetase (hGGPPS).	DE
PA	(HART/) HART E A.		XX
XX			XX
PI	Matsuda SPT, Hart EA;		KW
XX			XX
DR	WPI; 2004-373921/35.		OS
XX			XX
XX	New unicellular organisms comprising exogenous nucleic acids encoding a		PN
PT	geranylgeranyl pyrophosphate and a diterpene synthase, useful for		XX
PT	producing diterpenes and diterpene precursors.		XX
XX			PD
PS	Claim 30; SEQ ID NO 34; 38pp; English.		XX
XX			XX
CC	The invention relates to a unicellular organism for producing a diterpene		PF
CC	or diterpene precursor comprising an exogenous nucleic acid sequence		XX
CC	encoding a geranylgeranyl pyrophosphate synthase under the control of a		PR
CC	promoter operable in the organism, and an exogenous nucleic acid sequence		XX
CC	encoding a diterpene synthase under the control of a promoter operable in		PA
CC	the organism. The invention also relates to methods of producing a		XX
CC	diterpene or diterpene precursor and a method of isolating a diterpene		PI
CC	synthase comprising growing several cells in the presence of a		XX
CC	polyaromatic resin to make a cell/resin mixture, where at least one of		DR
CC	the cells further comprises at least one isolated and purified nucleic		XX
CC	acid sequence of a yeast expression library, and the expression of the		DR
CC	nucleic acid sequence is regulated by an inducible promoter under		XX
CC	conditions where the expression is induced, filtering the cell/resin		XX
CC	mixture, extracting the cell/resin mixture with alcohol to produce an		XX
CC	organic eluent and analysing the organic eluent by a screening method		XX
CC	including chromatography and/or spectroscopy, to identify the nucleic		PS
CC	acid sequence encoding the diterpene synthase. The unicellular		XX
CC	microorganism is useful as a diterpene or diterpene precursor producing		XX
CC	system. Diterpenes, in plants, serve as defence toxins, volatile		CC
CC	defensive signals, pollinator attractants and photoprotectants. This		CC
CC	sequence represents a geranylgeranyl pyrophosphate synthase polypeptide		CC
CC	of the invention. Note: The sequence data for this patent did not form		CC
CC	part of the printed specification but was obtained in electronic format		CC
CC	from USPTO at seqdata.uspto.gov/sequence.html.		CC
XX			XX
SQ	Sequence 303 AA;		SQ
Query Match 34.4%; Score 591.5; DB 8; Length 303;			
Best Local Similarity 41.1%; Pred. No. 3e-47;			
Matches 113; Conservative 67; Mismatches 78; Indels 17; Gaps 5;			
QY	20	ESLISKPYNHILKPGKFNELNLIQVIRVMNLPKDQALVSOIVLHNSLLIDIED 79	
DB	14	EDILEFYTLLISQPKDRAKILISAPDLMLHVPKDVLCVINKLIGHNASHMIDVDQ 73	
QY	80	NAPLRGQTTSHLIFGVPSITINTANYMYFRAMQVLSQTTKPELYHNLIIFNEELINLH 139	
DB	74	DSLRLRGVPAHHIYGVQPIINTANYVIFLALQEVKMLNIP-----SMMQVCTEELINLH 128	
QY	140	RGQGLDIYWRDPELPIIPTQEMYNLMVNMKTGGLFRLTLRLMALSPSSHCHSLVFPFN 199	
DB	129	RGQGLIYWRDSL--TCPTTEEYIDVMNNKTSGLLRLAVELMQAESD---IDYTPLVN 183	
QY	200	LLGIIVQIRDDYLNKDFQMSSEKGAEDITEGKLSPPVHAINFTKQTEQHNEILR 259	
DB	184	IIGHVQVRDDYVNMQLSTSYNNKGFCEDETEGKFSFPIIHAIK-----KDPNSRQLLN 237	
QY	260	ILLRLTSDKDIKULKIQLIFEDTNSLAYTKNFNFQ 294	
DB	238	IISQKPTSIEVKVYALEVIR-KAGSFEYVREFLRQ 271	
RESULT 9			
AA68909			

Db 233 ILRQRTENIDIKKVCVHYLE-DVGSFEYTRNLKEL-----EAK-----AYKQIDAR 278

QY 320 NLHDELLIYIDHSEL 335

Db 279 GGNPELVALVKHLSKM 294

RESULT 10

ADJ10383

ID ADJ10383 standard; protein; 300 AA.

AC ADJ10383;

XX 17-JUN-2004 (first entry)

DT

DE Human geranylgeranyl diphosphate synthase 1 protein.

XX

XX human; geranylgeranyl diphosphate synthase 1; GGPS1;

KW geranylgeranyl pyrophosphate synthetase; GGPPS; ggppsase;

KW geranyltransferase; embryonic development; cell differentiation;

KW apoptosis; developmental; hyperproliferative disorder; cancer;

KW cytostatic.

XX

OS Homo sapiens.

XX

XX US2004005570-A1.

XX

XX 08-JAN-2004.

XX

XX 02-JUL-2002; 2002US-00189268.

XX

XX 02-JUL-2002; 2002US-00189268.

XX

XX (ISIS-) ISIS PHARM INC.

XX

XX Dean NM, Bennett CF, Dobie KW;

XX

XX WPI; 2004-201146/19.

XX

XX N-PSDB; ADJ10255, ADJ10255.

XX

XX New antisense oligonucleotides for modulating geranylgeranyl diphosphate

PT synthase 1 expression, useful for diagnosing, preventing or treating

PT conditions associated with the protein, e.g. cancer.

XX

XX Disclosure; Page 31-32; 76pp; English.

XX

XX This invention relates to a novel antisense compounds that modulate the

CC expression of human geranylgeranyl diphosphate synthase 1 (also known as

CC GGPS1, geranylgeranyl pyrophosphate synthetase, GGPPS, ggppsase and

CC geranyltransferase) and located on chromosome 1p43. Specifically, it

CC refers to compositions useful for inhibiting the expression of GGPS1,

CC which normally participates in embryonic development, cell

CC differentiation and stimulates apoptosis via caspase-3 activation. The

CC present invention describes antisense oligonucleotides that comprise at

CC least one modified sugar moiety, a 2'-O-methoxyethyl (2' MOE) and at

CC least one modified nucleobase, a 5-methylcytosine. Accordingly, these

CC compounds are useful for treating a disease or condition associated with

CC GGPS1 such as a developmental or hyperproliferative disorder (e.g.

CC cancer) that arise as a result of aberrant apoptosis. As such, these

CC compositions exhibit cytostatic activity and are useful for diagnostics,

CC prophylaxis, research reagents and various kits. This polypeptide

CC sequence is the geranylgeranyl diphosphate synthase 1 protein of the

CC invention. NOTE: This sequence is given as an embedded protein in the

CC sequence listing and is not further referred to in the specification.

XX

XX Sequence 300 AA;

XX

Query Match 33.9%; Score 583; DB 8; Length 300;

Best Local Similarity 40.5%; Pseq No. 1.9e-46;

Matches 128; Conservative 55; Mismatches 103; Indels 30; Gaps 8;

QY 20 ESLISKYPNHLKPGKNFRLNIVQINRVNMLPKDLAIQVSIQVELHNSLLIDIED 79

Db 9 QRILLEPKYLQLPEKQVRKLSQAFNHLKVPEDKLOIIEVTEMLHNASLLIDIED 68

QY 80 NAPLRGQTTSHLIFGVSTINTANTMYFRAMQLVSOITTKPELYHNLITITNEELINLH 139

Db 69 NSKLRRGFPVAHSIVGIPSVINSANYVFLGLEKV--LTLDHP---DAVKLFTRQLLELH 123

QY 140 RGQGLDIYWRDFPEIIPTEQMYLNNVNNKTKGGLFRLRLRLMEALSPSSHGHSHSVPTFN 199

Db 124 QGQGLDIYWRD--NYTCTPEEYKAMVLQKTKGGLFGLAVGLMLQLF---SDYKEDLKPLIN 178

QY 200 LLGIYQIRDDYLNKDFQMSSEKGFARDITEGKLSFPIVHALNFTTKGQTEQNEILR 259

Db 179 TLGLFFQIRDDYANLHSEYSEKSCFCEDETEGKFSFTTHAI-----WSRPESTQVQN 232

QY 260 ILLRTSDKDIKLIQILEFDTNLSLAYTKNFINQLVNMKNKNENKYLPLDASHSDTAT 319

Db 233 ILRQRTENIDIKKVCVHYLE-DVGSFEYTRNLKEL-----EAK-----AYKQIDAR 278

QY 320 NLHDELLIYIDHSEL 335

Db 279 GGNPELVALVKHLSKM 294

RESULT 11

ADM98626

ID ADM98626 standard; protein; 300 AA.

XX

XX ADM98626;

XX

XX 01-JUL-2004 (first entry)

XX

XX Geranylgeranyl pyrophosphate synthase polypeptide #25.

XX

XX Geranylgeranyl pyrophosphate synthase; diterpene; diterpene precursor;

KW diterpene synthase; defence toxin; volatile defensive signal;

KW pollinator attractant; photoprotectant; enzyme.

XX

XX Homo sapiens.

XX

XX US2004072323-A1.

XX

XX 15-APR-2004.

XX

XX 07-JAN-2002; 2002US-00041018.

XX

XX 05-JAN-2001; 2001US-0259880P.

XX

XX (MATS/) MATSUDA S P T.

XX (HART/) HART E A.

XX

XX Matsuda SPT, Hart EA;

XX

XX WPI; 2004-373921/35.

XX

XX New unicellular organisms comprising exogenous nucleic acids encoding a

PT geranylgeranyl pyrophosphate and a diterpene synthase, useful for

PT producing diterpenes and diterpene precursors.

XX

XX Claim 30; SEQ ID NO 46; 38pp; English.

XX

XX The invention relates to a unicellular organism for producing a diterpene

CC or diterpene precursor comprising an exogenous nucleic acid sequence

CC encoding a geranylgeranyl pyrophosphate synthase under the control of a

CC promoter operable in the organism, and an exogenous nucleic acid sequence

CC encoding a diterpene synthase under the control of a promoter operable in

CC the organism. The invention also relates to methods of producing a

CC diterpene or diterpene precursor and a method of isolating a diterpene

CC synthase comprising growing several cells in the presence of a

CC polyaromatic resin to make a cell/resin mixture, where at least one of

CC the cells further comprises at least one isolated and purified nucleic

CC acid sequence of a yeast expression library, and the expression of the

CC nucleic acid sequence is regulated by an inducible promoter under

CC conditions where the expression is induced, filtering the cell/resin
 CC mixture, extracting the cell/resin mixture with alcohol to produce an
 CC organic eluent and analysing the organic eluent by a screening method
 CC including chromatography and/or spectroscopy, to identify the nucleic
 CC acid sequence encoding the diterpene synthase. The unicellular
 CC microorganism is useful as a diterpene or diterpene precursor producing
 CC system. Diterpenes, in plants, serve as defence toxins, volatile
 CC defensive signals, pollinator attractants and photoprotectants. This
 CC sequence represents a geranylgeranyl pyrophosphate synthase polypeptide
 CC of the invention. Note: The sequence data for this patent did not form
 CC part of the printed specification but was obtained in electronic format
 CC from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 300 AA;

Query Match 33.9%; Score 583; DB 8; Length 300;
 Best Local Similarity 40.5%; Pred. No. 1.9e-46;
 Matches 128; Conservative 55; Mismatches 103; Indels 30; Gaps 8;

QY 20 ESLISKPNHLLKPKNFRLNLIQVQINRVMNLPKDOLAIVSQIVVELLHNSLLIDDDIED 79
 DB 9 QRLLEPKYLLQOLPKQVTKLSQAFNHWLKVPEKQGLIIIEVTMLHNSLLIDDDIED 68
 QY 80 NAPLRGQTTSHLIFGVPTINTANYMYPRAMQVLSQLTKEPLVHNLIITIFNEELINLH 139
 DB 69 NSKLRRGFPVAHSIYGIPSVINSANVYFLGLEKV--LTLDPH---DAVKLFRQLLELH 123
 QY 140 RQGLDIYWRDPFLPELIPTQEMLYNMVNMTGGLFRLTLRLMEALSPSHHSHLVFFIN 199
 DB 124 QQGLDIYWRD--NYTCPTTEEYKAMVLQKTGGLFGLAVGLMQLF---SDYKEDLKPLIN 178
 QY 200 ILGITVQIRDDVNLKDFQMSSEKGFPAEDITGKLSFPVTHALNFTKTKGQTEQHNEILR 259
 DB 179 TLGLFFQIRDDYANLHSHKEYSENKSCEDLTGKFSFPTIHAI-----WSRPESTQVQN 232
 QY 260 ILLTSDXDKIKLITQILEFDTNSLAYTKNFINQVNMKNNDENKYLPLDLASHSDTAT 319
 DB 233 ILRQRTENIDIKKYCVHYLE-DVGSPEYTRNTLTKEL-----EAK-----AYKQIDAR 278
 QY 320 NLHDELLYIDHSEL 335
 DB 279 GGNPELVALVKHLSKM 294

RESULT 12
 AAB58276
 ID AAB58276 standard; protein; 304 AA.

XX AAB58276;

XX 14-MAR-2001 (first entry)

XX Lung cancer associated polypeptide sequence SEQ ID 614.

XX Human; lung cancer associated protein; neuroprotective; cytostatic;
 XX cardioactive; immunomodulatory; muscular active; vulnary;
 XX gastrointestinal; nephrotropic; antiinfective; gynecological;
 XX antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 XX proliferative disorder; wound healing; infectious disease.

XX Homo sapiens.

XX WO200055180-A2.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US005918.

XX 12-MAR-1999; 99US-0124270P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX

PI Ruben SM;

XX WPI: 2000-587514/55.
 DR N-PSDB; AAF18152.

XX Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer.

XX Claim 11; Page 1108-1109; 1425pp; English.

XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardioactive;
 CC immunomodulatory; muscular active general; vulnary; gastrointestinal
 CC general; nephrotropic; antiinfective; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the protein
 CC or polynucleotide sequences. The lung cancer associated polynucleotide
 CC sequences may be used for detection of lung cancer, chromosome
 CC identification, as chromosome markers, and for numerous other diagnostic
 CC or research purposes. The proteins may be used to treat disorders such as
 CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
 CC cardiovascular, renal, and proliferative disorders. The proteins may also
 CC be used in the treatment of wounds and infectious diseases.
 CC Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are
 CC used in the course of the invention for the identification and
 CC characterisation of the polynucleotide and protein sequences

XX Sequence 304 AA;

Query Match 33.5%; Score 577; DB 3; Length 304;
 Best Local Similarity 39.6%; Pred. No. 7e-46;
 Matches 129; Conservative 58; Mismatches 107; Indels 32; Gaps 9;

QY 12 DPWSSQN--ESLISKPNHLLKPKNFRLNLIQVQINRVMNLPKDOLAIVSQIVELH 69
 DB 3 NPMEKTQETVQRIILEPKYLLQOLPKQVTKLSQAFNHWLKVPEKQGLIIIEVTMLH 62

QY 70 SSLLIDDDIEDNAPLRGQTTSHLIFGVPTINTANYMYPRAMQVLSQLTKEPLVHNLI 129
 DB 63 ASLLIDDDIEDNAPLRGQTTSHLIFGVPTINTANYMYPRAMQVLSQLTKEPLVHNLI 117

QY 130 IFNEELINLHRGQGLDIYWRDPFLPELIPTQEMLYNMVNMTGGLFRLTLRLMEALSPSH 189
 DB 118 LFRQLLELHQGGLDIYWRD--NYTCPTTEEYKAMVLQKTGGLFGLAVGLMQLF---SD 172

QY 190 HGSLVFPINLGIYQIRDDVNLKDFQMSSEKGFPAEDITGKLSFPVTHALNFTKTKG 249
 DB 173 YKEDLKPLNLTGLFPQIRDDYANLHSHKEYSENKSCEDLTGKFSFPTIHAI-----W 226

QY 250 QTEQHNEILRIILLTSDXDKIKLITQILEFDTNSLAYTKNFINQVNMKNNDENKYL 309
 DB 227 SRXESTQVQNILRQRTENIDIKKYCVHYLE-DVGSPEYTRNTLTKEL-----EAK--- 274

QY 310 DLASHSDTATNLHDELLYIDHSEL 335

DB 275 --AYKQIDARGGNPELVALVKHLSKM 298

RESULT 13

AAW59753

ID AAW59753 standard; protein; 300 AA.

XX AAW59753;

XX 12-OCT-1998 (first entry)

XX Human geranyl pyrophosphate synthase.

XX Human; geranyl pyrophosphate synthase; hGPGS; antagonist;

XX agonist; ss.

Result No.	Query %		DB	ID	Description
	Score	Match Length			
1	1720	100.0	335	2	US-08-761-344-2
2	574	33.4	300	1	US-08-469-668-2
3	574	33.4	300	2	US-09-038-596-2
4	574	33.4	300	5	PCT-US95-00421-2
5	538	31.3	376	3	US-09-091-725-15
6	242.5	14.1	307	4	US-09-543-681A-5908
7	233	13.5	341	4	US-09-252-991A-2148
8	232	13.5	165	4	US-09-248-796A-17985
9	228.5	13.3	330	1	US-08-410-167A-4
10	228.5	13.3	330	2	US-08-898-560-1
11	227	13.2	342	4	US-09-489-039A-13653
12	225.5	13.1	325	3	US-09-217-609A-2
13	225.5	13.1	325	3	US-08-873-235B-2
14	225.5	13.1	330	3	US-09-103-126-1
15	221.5	12.9	340	4	US-09-543-681A-5778
16	210.5	12.2	333	3	US-09-025-819-29
17	210.5	12.2	333	3	US-09-808-126-29
18	210.5	12.2	333	4	US-09-803-951-29
19	206.5	12.0	162	4	US-09-248-796A-17986
20	204	11.9	353	4	US-09-328-352-4226
21	189	11.0	335	4	US-09-673-018-2
22	188	10.9	380	4	US-09-107-532A-4570
23	186.5	10.8	325	3	US-09-009-895-2
24	174.5	10.1	319	4	US-09-710-279-1040
25	174.5	10.1	357	3	US-09-134-001C-3250
26	171.5	10.0	297	4	US-09-934-903-14
27	171.5	10.0	297	4	US-09-934-868-72

Db 61 SQIVYELHNSLLDDIEDNAPLRRGQTSLSLIFGVPSNTANTMYFRAMQLVSLITK 120
Qy 121 BPLVHNLITIFNEELINLHROGLDIYWRDFELPELIPTQEMYLNNMKNKGTGLFRLTLRL 180
Db 121 BPLVHNLITIFNEELINLHROGLDIYWRDFELPELIPTQEMYLNNMKNKGTGLFRLTLRL 180
Qy 181 MEALSPSSHGHGSLVFPFINLLGIYQIRDDYVNLKDFQMSSEKGFPAEDITGKLSFPPIVH 240
Db 181 MEALSPSSHGHGSLVFPFINLLGIYQIRDDYVNLKDFQMSSEKGFPAEDITGKLSFPPIVH 240
Qy 241 ALNFTKTKGTEQHNHRIILRLPSDKDIKLIQILEFDNSLAYTKNFNLQVNMVNIK 300
Db 241 ALNFTKTKGTEQHNHRIILRLPSDKDIKLIQILEFDNSLAYTKNFNLQVNMVNIK 300
Qy 301 NDENKYLPLDASHSDTATNLHDELLYIIDHLSL 335
Db 301 NDENKYLPLDASHSDTATNLHDELLYIIDHLSL 335

RESULT 2
US-08-469-665-2
; Sequence 2, Application US/08469665
; Patent No. 5786193
; GENERAL INFORMATION:
; APPLICANT: GREENE, ET AL.
; TITLE OF INVENTION: Human Geranylgeranyl Pyrophosphate Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,665
; FILING DATE: June 6, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/00421
; FILING DATE: 11 JAN 95
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-377
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 300 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-469-665-2

Query Match 33.4%; Score 574; DB 1; Length 300;
Best Local Similarity 40.2%; Pred. No. 7.4e-47;
Matches 127; Conservative 55; Mismatches 104; Indels 30; Gaps 8;
Qy 20 ESLISKPYNHILLPGKNFRNLIVQINRVNLPKDLAIVSQIVELHNSLLDDIED 79
Db 9 QRILLEPKYLLQPLPGQVTKLSQAFNHLKVPEDKLIIEVTEMLHNASLLDDIED 68
Qy 80 NAPLRRGQTSLSLIFGVPSNTANTMYFRAMQLVSLITTKPEPLYHNLITIFNEELINLH 139

Db 69 NSKLRRGFVVAHSYIGIPSVINSANYVFLGLEKV--LTLDP--DAVKLFTQLLELH 123
Qy 140 RQGLDIYWRDFELPELIPTQEMYLNNMKNKGTGLFRLTLRLMEALSPSSHGHGSLVFPIN 199
Db 124 QGQGLDIYWR--NYTCTEETBEYKAWVLQKTGGFLGAVGLNQLF--SDYKEDLKPLN 178
Qy 200 LLGIYQIRDDYVNLKDFQMSSEKGFPAEDITGKLSFPPIVHALNFTKTKGTEQHNHRIIL 259
Db 179 TLGLFFQIRDDYVNLHSEYSENKSLGDLTEGKFSFTIHAI-----WSRSESTQVQN 232
Qy 260 ILLRTSDKDIKLIQILEFDNSLAYTKNFNLQVNMKNKDNENKYLPLDASHSDTAT 319
Db 233 ILKQRTENIDIKKYCVHYLE-DVSGSEYTRNTLKL-----EAK-----AYKQIDAR 278
Qy 320 NLHDELLYIIDHLSL 335
Db 279 GGNPELVALVKHLSKN 294

RESULT 3
US-09-038-596-2
; Sequence 2, Application US/09038596
; Patent No. 5928924
; GENERAL INFORMATION:
; APPLICANT: GREENE, ET AL.
; TITLE OF INVENTION: Human Geranylgeranyl Pyrophosphate Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/038,596
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,665
; FILING DATE: June 6, 1995
; APPLICATION NUMBER: PCT/US95/00421
; FILING DATE: 11 JAN 95
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-377
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 300 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-09-038-596-2

Query Match 33.4%; Score 574; DB 2; Length 300;
Best Local Similarity 40.2%; Pred. No. 7.4e-47;
Matches 127; Conservative 55; Mismatches 104; Indels 30; Gaps 8;
Qy 20 ESLISKPYNHILLPGKNFRNLIVQINRVNLPKDLAIVSQIVELHNSLLDDIED 79
Db 9 QRILLEPKYLLQPLPGQVTKLSQAFNHLKVPEDKLIIEVTEMLHNASLLDDIED 68

Qy	80	NAPLRRGQTTSHLIFGVPSTINTANYMYFRAMQLYSQLTKEPLVHNLTITIFNEBLINLH	139
Db	69	NSKLRRGFPAHSYIGIPSVINSANYVFLGLEKV--LTLDHP--DAVKLETRQLLELH	123
Qy	140	RGQGLDIYWRDPELPIPTOEMYLMVMNKNKTGGFLRTLRLMEALSPSHHGHSVLPFTN	199
Db	124	QGQGLDIYWRD--NVTCTPEEYKAWVLQKTGGLFGLAVGLMQLF--SDYKEDLKPLN	178
Qy	200	LGHIYIORDYLNLRKQFQMSKGFABDITGKLSFPFIVHALNPTKKGQTEQRNEILR	259
Db	179	TLGLFFOIRDDYANLHRSKEYSENKSGDLTEGFSFPTIHAI-----WRSRSESTQVQN	232
Qy	260	ILLRTSDKDIKLKLIQLLEPDTNSLAYTKYFINQLVNMINKDNENKYLPLDLASHSDTAT	319
Db	233	ILRQRTENIDIKKVCVHYLE-DVCSGEYTRNTLKEI-----EAK-----AYKQIDAR	278
Qy	320	NLHDELLYIIDHLSBEL	335
Db	279	GENPELVALVKHLSKM	294

```

1  RESULT 5
2  US-09-091-725-15
3  ; Sequence 15, Application US/09091725
4  ; Patent No. 6329141
5  ; GENERAL INFORMATION:
6  ;
7  ; APPLICANT:
8  ; TITLE OF INVENTION: Improved methods for transforming Phaffia
9  ; TITLE OF INVENTION: and recombinant DNA for use therein
10 ;
11 ; NUMBER OF SEQUENCES: 51
12 ; CORRESPONDENCE ADDRESS:
13 ; ADDRESSEE: Morrison & Foerster lip
14 ; STREET: 2000 Pennsylvania Avenue, N.W.
15 ; CITY: Washington
16 ; STATE: DC
17 ; COUNTRY: United States of America
18 ; ZIP: 20006-1898
19 ;
20 ; COMPUTER READABLE FORM:
21 ; MEDIUM TYPE: Floppy disk
22 ; COMPUTER: IBM PC compatible
23 ; OPERATING SYSTEM: PC-DOS/MS-DOS
24 ; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
25 ; CURRENT APPLICATION DATA:
26 ; APPLICATION NUMBER: US/09/091,725
27 ; FILING DATE: 23-DEC-1996
28 ; CLASSIFICATION: 435
29 ; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: EP 95203620.0
 FILING DATE: 22-DEC-1995
 APPLICATION NUMBER: EP 96200943.7
 FILING DATE: 11-APR-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: E. Victor Donahue
 REGISTRATION NUMBER: 35,492
 INFORMATION FOR SEQ ID NO: 15:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 376 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-091-725-15

	Query Match	31.3%;	Score 538;	DB 3;	Length 376;	
	Best Local Similarity	31.1%;	Pred. No. 2.9e-43;			
	Matches 117;	Conservative 79;	Mismatches 110;	Indels 70;	Gaps 7;	
OY	15 WSSQNESLISKPNHILKPGKNFRNLIVQNVRNLPDQQAIVSQVIELLNSSLII	74				
dDb	:: :: :: :: :: :: :: :: :: :: :: ::					
	14 FTPODDI VLEPYHLGKNGPKGEIRSQLFEAFNYWLDVKEDLEVIGNVVGLHTASLLM	73				
OY	75 DDIEDNAPIRRQTTHSLIFGVGPSTINTANYMYFRAMLQV-----	114				
dB	74 DVEDSSVLRRGSVAHYLTGIPTOINTANYVFAYQEIKFLRPITPMWPVIPSPSASI	133				

115 -----SOLTKRPLVHNLITIFNEELINLHRCGLDIYVRD 150
134 QSSVSSASSSSASSENGSTPNSQIPFSDTY--LDKVTIDEMLSLHRCGLLELFRD 191
151 FLPEIIPQEMVLMNMVNTGLRLTLRLMEALSPSSHHGSLVFPFNLGLIYQIRDD 210
192 SL--TCPSEBYVVMVKGKTGGLFRIAVRLMA--KSECIDFQVAVNLISYIFQIRDD 246
211 YLNKDFQMSSEKGFPAEDITEKSPPIVHALNFTKQTEQHNEIRILLRTSDKDI 270
247 YMNLOSSEYAHKNFAEDITEKGFSPFTIHSIHANPSS-----RLVINTLQKKSTSEI 300
271 KKLQILEFDNLSLAYTKNFNLVNMKNKND-----NENKYLPLDASHSDTAT 319
301 LHCVNYRTTETHSFTEVLTLSGALRELRLQGEFAFANSRMDLGDVDSGRGCK 360
320 NLHDELLYIIDLSEL 335
361 NV--KLEALLKLAADI 374

RESULT 6
US-09-543-681A-5908
; Sequence 5908, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5908
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5908

Query Match 14.1%; Score 242.5; DB 4; Length 307;
Best Local Similarity 25.2%; Pred. No. 4.5e-15;
Matches 77; Conservative 67; Mismatches 132; Indels 29; Gaps 8;

QY 4 KIDELINNDPVWSSQNESLISKPNHLLKPGKFRNLIVQINRVNMLPKDQIAIVSQI 63
Db 21 RVDEALNQALQTLFPFSDMPLSQAMRYGALGGKRLRPFLVYAVGEMFNVPVANLDVFAA 80
QY 64 VELLHNSLLIDIE--DNAPLRGQTTSHLIFGVPSTINTANYMYFRAMQVLSLTKE 121
Db 81 IECIHAYSLHDDLPAWMDDLRGRKPTCHIEFGEANAILAGALQTLAEFIILAKAMPD 140
QY 122 PLYHNLITFNE-----ELINLHRCGLDIYVRDFLPEIIPQEMVLMNMVNTKGLFRL 176
Db 141 VATADRVMATKATASGLAGMCGQALDDAEDKSDIVALEKHL---HKTGALIRA 196
QY 177 TLRMEALSPSSHHGSLV-----FINLGIYQIRDDYLNKDFQMSSEKGFPAEDITEG 232
Db 197 AVRL--GALS--AGQKGDVLPALDKYAHSIGLAFQVQDDILDVIGSTEETGKRGSDQEAG 254
QY 233 KLSFPPIVHALNFTKQTEQHNEIRILLRTSDKDIKLQILEFDNLSLAYTKNFI 292
Db 255 KSTYPALLGLAQAKQKQAO--ELYNEALD-----ALAFLEYDYDITSLKQLANFI 302
QY 293 NOLVN 297
Db 303 VERKN 307

RESULT 7
US-09-252-991A-21448

; Sequence 21448, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21448
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21448

Query Match 13.5%; Score 233; DB 4; Length 341;
Best Local Similarity 21.5%; Pred. No. 4.3e-14;
Matches 65; Conservative 81; Mismatches 135; Indels 22; Gaps 6;

QY 22 LISKPNHLLKPGKFRNLIVQINRVNMLPKDQIAIVSQIIVHLLHNSLLIDIEDNA 81
Db 50 LVEKIDGVIISAGGRRLRPFLVLLAGKTIGYKGGDCLLAATIFRLTSTLLHDDVDAS 109
QY 82 FLRRQTTSHLIFGVPSTINTANYMYFRAMQVLSLTKEPLYHNLITIFNEELINLHRCG 141
Db 110 GLRRGRSTANALWGNAPSVLVGDFLYARSFEMVELGSMW-----VVRILSQATRVIAEG 164
QY 142 QCLDIYVRDFLPEIIPQEMVLMNMVNTKGLFRLTLRLMEAL--SPSSHHGSLVFPFNL 200
Db 165 EVLQL---SKVRDASTTEETVMEVIRGTAMLFEAATSTHSAALCOAGEHQSEALRRFGDY 221
QY 201 LGIYQIRDDYLNKDFQMSSEKGFPAEDITEKSPPIVHALNFTKQTEQHNEIRIL 260
Db 222 LGIAFQVLDLDYRGDAATLGKNVGGDLAEGKPTPLIVTM-----RDTBEQAALVRK 276
QY 261 LLRLTSDKDIK-----LKLQILEFDNLSLAYTKNFNLVNMKNKNDNENKYLPLDASHS 315
Db 277 ALCQGGSQLESVCAAVEAAGALDYTA---LADYAAARAIACLDLTPDNEYRSALVELS 333
QY 316 DTA 318
Db 334 EFA 336

RESULT 8
US-09-248-796A-17985
; Sequence 17985, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17985
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-17985

Query Match 13.5%; Score 232; DB 4; Length 165;
Best Local Similarity 36.8%; Pred. No. 1.9e-14;

Matches 67; Conservative 30; Mismatches 57; Indels 28; Gaps 8;

```

QY 162 YLNMVANKTGLFRLTLRLMEALSPSHHGHSLVPPINILGIIYQIRDDYLNKDFQMS 221
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 4 YLEMKDKTGGFRLAUKLILLYSDVTN-DQLISLANLWGLIYQVRDYLNLVDKYS 62
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
QY 222 EKG-PAEDITEKLSFPIVHALNFTTKGTQTEQHNEILRLILRTSDKD-----IKLK 273
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 63 MKGTTCEDLIEGLSLFILHCLRTTK-----DSPVHKILYDSDSSDRVSQNSLIDLS 115
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
QY 274 LIQILEFTNLSLAYTKNFQNLVNMKNDNENKYLPLDLASHSDTATNLHDELLYIIDHLS 333
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 116 L-SFMNKSLSLEYTLNLKLVLEKKLR-QLILKY-PELENSA-----LLKIVERIC 163
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
QY 334 EL 335
   :|
Db 164 DL 165

```

```

RESULT 9
US-08-410-167A-4
; Sequence 4, Application US/08410167A
; Patent No. 5773273
; GENERAL INFORMATION:
; APPLICANT: TOKUZO NISHINO, Shinichi OHNUMA, Manabu SUZUKI,
; APPLICANT: Chikara OHTO, Chika ASADA, Yuka HIGUCHI, Yoshie TAKEUCHI
; TITLE OF INVENTION: Geranylgeranyl-Diphosphate Synthase and DNA
; TITLE OF INVENTION: Coding Therefor
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: One Broadway
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3+ Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,167A
; FILING DATE: 24-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-53804
; FILING DATE: 24-MAR-1994
; APPLICATION NUMBER: JP 6-315572
; FILING DATE: 25-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Edward W. Greason
; REGISTRATION NUMBER: 18,918
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)425-7200
; TELEFAX: (212)425-5288
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Sulfolobus acidocaldarius
; STRAIN: ATCC 33909
US-08-410-167A-4

```

Query Match 13.3%; Score 228.5; DB 1; Length 330;
 Best Local Similarity 25.8%; Pred. No. 1.1e-13;
 Matches 83; Conservative 65; Mismatches 151; Indels 23; Gaps 10;

```

QY 6 DELINN-DPVMWSQNESLISKPN--HLLKPGKNFRNLNLVQINRVNMLPKDQIAIVS 61
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 9 NEIVSNVNDIISKYISGDVPKLYEASVYHFTSGGKRLRPLILTISSDLFGGQERAYYAG 68
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
QY 62 QIVELLHNSLLIIDDIEDNAPLRGOTTSHLIFGVSPSTINTANYMYFRAMQLVSQLTKE 121
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 69 AAIEVLHFTLVHDDIMQDNIRRLGPIVHVVKYGLPLAILAGDLLHAKAFQLLTQALRGL 128
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
QY 122 P--LYHNLIITPNEELINLHROGLDIYWRDPLPHIIPTOEMYLNMVNMKTGGLFRLTLR 179
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 129 PSETTIKAFDIFTRSIISIIISGQAVDMFEFDRID--IKEQE-YLDWISRKTAALFSASS 185
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
QY 180 LMEALSPSSHGHSLV-PFINLGIYIQRDDYLNKDFQMSSEKGFPAEDITEGKLSFPI 238
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 186 IGAITAGANDNDVRLMSDFGTNLGIAFQIVDDILGLTADKELGKPFVSDIREGKTTIV 245
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
QY 239 VHALNFTTKGTQTEQHNEILRLILRTSDKDILKLIQILEFTNLSLAYT-----KNFI 292
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 246 IKTLELCK---EDEKKIVLKAIGNKSASKELMSSADIIKKYSLDYAVNLAETYYKNAI 301
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
QY 293 NQLVNM--KNDNENKYLPLDLA 312
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 302 DSL-NQVSSKSDIPGKALKYLA 322
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :

```

```

RESULT 10
US-08-898-560-1
; Sequence 1, Application US/08898560
; Patent No. 5935832
; GENERAL INFORMATION:
; APPLICANT: HIROYUKI NAKANE, Chikara OHTO, Shinichi OHNUMA,
; APPLICANT: Kazutake HIROOKA, Tokuzo NISHINO
; TITLE OF INVENTION: Farnesyl Diphosphate Synthase
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: One Broadway
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3+ Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/898,560
; FILING DATE: Concurrent Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-213211
; FILING DATE: 24-JUL-96
; ATTORNEY/AGENT INFORMATION:
; NAME: Edward W. Greason
; REGISTRATION NUMBER: 18,918
; REFERENCE/DOCKET NUMBER: 77670/495
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)425-7200
; TELEFAX: (212)425-5288
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Sulfolobus acidocaldarius
; STRAIN: ATCC 33909
; FEATURE:
; NAME/KEY: Asp-rich domain
; LOCATION: 82-86
US-08-898-560-1

```


QY 295 LVNMIKNDNENKYLDEL--ASHSDTATNLHDELL 326
Db 288 LLSRKYGDKAKYHLSQLQDSNLIKDYLEIHEKML 321

RESULT 13

US-08-873-235B-2
; Sequence 2, Application US/08873235B
; Patent No. 6174715
; GENERAL INFORMATION:
; APPLICANT: MURAMATSU, Masayoshi
; APPLICANT: KOIKE, Ayumi
; APPLICANT: OGURA, Kyoza
; APPLICANT: KOYAMA, Tanetoshi
; APPLICANT: SHIMIZU, Naoto
; APPLICANT: CHO, Yenwin
; TITLE OF INVENTION: Prenyl Diphosphate Synthetase Genes
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, NW - Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
; SOFTWARE: Wordperfect 6.1 for Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/873,235B
; FILING DATE: 11-Jun-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 154441/1996
; FILING DATE: 14-Jun-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: TOFFENETTI, Judith L.
; REGISTRATION NUMBER: 39,048
; REFERENCE/DOCKET NUMBER: 10235/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-429-1776
; TELEFAX: 202-429-0796
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-873-235B-2

Query Match 13.1%; Score 225.5; DB 3; Length 325;
Best Local Similarity 24.3%; Pred. No. 2.1e-13;
Matches 81; Conservative 72; Mismatches 144; Indels 37; Gaps 11;
QY 1 MEAKIDELINNDPWSQSONESLISKPNHLLKPKGNRLNLIVQINRVMNLPKQDLAIV 60
Db 17 VEKRLVECTQSD-----SETINKAAHHLSSGGRKVRPMFVLLSGFLMDTQKDLIRT 69
QY 61 SQIVELLHNSLLDIEDNAPLRGGOTTSHLIFGVPSSTINTANYMYFRAMQLVSQLTTK 120
Db 70 AVSLELVHMASLVHDDYIDNSDMERGNTSVHIAFDKDTARTGHFLARALQNIATINNS 129
QY 121 EPLVHNLITIFNEELINLHRGQGLDIYWRDFLEIPIPTQEMYLNVMNKTGGFLRLTLR 180
Db 130 K--FHQ---IFSKTILEVCFGEFDQADRENYPVSEFTA---YLRRINRKTAILIEASCHL 181
QY 181 MEALS-----PSSHHGSHSLV--PFINLGIYQIRDDYINLKFQMSSEKGFADITEGKLSFPI 235
Db 182 -GALSSQLDEQSYH---IKQFGHIGMSYQIIDDILDTYSDEATLGKPVGSDIRNGHIT 237
QY 236 FPIVHALNFTTKGQTEQHNEILRIILLRTSDKDIKLLI-QILEFTDNTSLAYTKKFINQ 294

Db 238 YPLMAAI--ANLKEQDDDDKLEAVVKHLTSTSDDEVYQIVSQVKYGLEPA-----E 287
QY 295 LVNMIKNDNENKYLDEL--ASHSDTATNLHDELL 326
Db 288 LLSRKYGDKAKYHLSQLQDSNLIKDYLEIHEKML 321

RESULT 14

US-09-101-126-1
; Sequence 1, Application US/09101126
; Patent No. 6316216
; GENERAL INFORMATION:
; APPLICANT: OHTO, CHIKARA
; APPLICANT: NAKANE, HIROYUKI
; APPLICANT: NISHINO, TOKUZO
; APPLICANT: OHNUMA, SHINTACHI
; APPLICANT: HIROOKA, KAZUTAKE
; TITLE OF INVENTION: MUTATED PRENYL DIPHOSPHATE SYNTHASES
; FILE REFERENCE: 77670/566
; CURRENT APPLICATION NUMBER: US/09/101,126
; CURRENT FILING DATE: 1999-04-27
; EARLIER APPLICATION NUMBER: PCT/JF97/03921
; EARLIER FILING DATE: 1997-10-29
; EARLIER APPLICATION NUMBER: JP 8-307506
; EARLIER FILING DATE: 1996-11-05
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 1
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Sulfolobus acidocaldarius
; FEATURE:
; OTHER INFORMATION: 82-86 is an Asp-rich domain
US-09-101-126-1

Query Match 13.1%; Score 225.5; DB 3; Length 330;
Best Local Similarity 25.8%; Pred. No. 2.1e-13;
Matches 83; Conservative 64; Mismatches 152; Indels 23; Gaps 10;
QY 6 DELINN-DPVWSSQNESLISKPNY--HILLKPKGNRLNLIVQINRVMNLPKQDLAIVS 61
Db 9 NEIVNSVNDIISKVYISGVDPKLYEASVHLFTSGKRLRPILITSSDLFGQGRGAYYAG 68
QY 62 QIVELLHNSLLDIEDNAPLRGGOTTSHLIFGVPSSTINTANYMYFRAMQLVSQLTTKE 121
Db 69 AAIEVLHTFTLVHDDIMDQDNIRRGPLTVHVKYGLPLAILAGDLLHAKAFQLLTQALRGL 128
QY 122 P--LYHNLIITFNEELINLHRGQGLDIYWRDFLEIPIPTQEMYLNVMNKTGGFLRLTLR 179
Db 129 PSETIIKAFDIFTRSIISSEGOAVDMFEEDRID--IKEQE-YLDMISRTAALFSASS 185
QY 180 LMEALSPSSHGHSHSLV--PFINLGIYQIRDDYINLKFQMSSEKGFADITEGKLSFPI 238
Db 186 IGALIAAGANDNDVRLMSDFGTNIGIAFQIVDDIIGLTADEKELGKPVFSDIRGKKTILV 245
QY 239 VHALNFTTKGQTEQHNEILRIILLRTSDKDIKLLIQLIEFTDNTSLAYT-----KNFT 292
Db 246 IKTLELCK----EDEKKIVLKALGNKSKASKEELMSADIKKYSLDYAVNLAEEKYKNAI 301
QY 293 NQLVNMII--KNDNENKYLPLDA 312
Db 302 DSL-NOVSSKSDIPGKALKYLA 322

RESULT 15

US-09-543-681A-5778
; Sequence 5778, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

```
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5778
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5778

Query Match      12.9%; Score 221.5; DB 4; Length 340;
Best Local Similarity 23.1%; Pred. No. 5.4e-13;
Matches 68; Conservative 71; Mismatches 114; Indels 41; Gaps 8;

Qy      11 NDPVSSQNE--SLISKPYNHILLKPGKNFRNLIVQINRVNMLPKDQIAIVSQIVELLH 68
Db      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Qy      35 NDAILYQLNSDVSLINQLGYIISGGKRIKRPMTIAVLTAQALDYQCDKHISVAALIEFIH 94
Db      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Qy      69 NSSLIIDIEDNAPLRGQTTHLIFGVPSNTANTANYMYFRAMQLVSQLTTKEPIYHNL 128
Db      : ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Qy      95 TATLHDDVDVDESDMRGKQTANAVFGNAASVLVGVDFIYTRSFQMTDLDLSNRVL--KLM 152
Db      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Qy      129 T-----IFNEELINLHRCQGLDIYWRDPLPEIIPQEMVLYNNMKNKGTGLFRLTLRLMEA 183
Db      : ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Qy      153 SSATNVIAGEVQLMNCNDPDI-----TEDDYMQVIYSKTA-----RLFEA 194
Db      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Qy      184 LSPSS-----HGHSLVPPFINLLGIYQIRDDYLNKDPQMSSEKGFADITEGKLS 235
Db      : ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Qy      195 ASHASAILCGATPEQEKAFQDYGRVILGTAFQLIDLDLDYDADNTQLGKNTGDDLDGKPT 254
Db      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Qy      236 PPIVHALNFTKGTQEQHNEILRILLRTSDKDIKLIQILEFDTSLAYTK 289
Db      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Qy      255 LPLHAMHH-----GNETESQLIROAIEKNGRHLDDTVLTKMK--QCGSLEYTR 302
Db      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
```

Search completed: October 17, 2004, 08:56:04
Job time : 29.4406 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 23:38:29 ; Search time 729.274 Seconds
(without alignments)
11005.915 Million cell updates/sec

Title: US-10-041-018-1

Perfect score: 1569

Sequence: 1 aatattacatagatagatag.....aatgccgtaattcgcga 1569

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 3403857 seqs, 2557783690 residues

Total number of hits satisfying chosen parameters: 6807714

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:**
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq**
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq**
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq**
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq**
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq**
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq**
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq**
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq**
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq**
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq**
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq**
12: /cgn2_6/ptodata/1/pubpna/US09D_PUBCOMB.seq**
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq**
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq**
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq**
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq**
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq**
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq**
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq**
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1569	100.0	1569	16	US-10-041-018-1
2	1008	64.2	1008	15	US-10-369-493-46181
3	1008	64.2	1008	16	US-10-450-941-5
4	114	7.3	1395	16	US-10-189-268-4
5	114	7.3	1395	10	US-09-814-353-21762
6	114	7.3	2534	16	US-10-041-018-3
7	113.6	7.2	2529	9	US-09-925-302-171
8	113.6	7.2	2529	10	US-09-925-302-171
9	102.6	6.5	1044	15	US-10-369-493-36092
10	90.6	5.8	1284	15	US-10-369-493-27908
11	87.6	5.6	51001	16	US-10-189-268-11
12	84.2	5.4	2222	16	US-10-041-018-21
13	70.6	4.5	1882	16	US-10-001-192A-14

C 14	68.8	4.4	554	11	US-09-969-034-1324	Sequence 1324, Ap
C 15	68.4	4.4	716	11	US-09-969-034-1324	Sequence 1324, Ap
C 16	65.6	4.2	593	11	US-09-969-034-1488	Sequence 1488, Ap
C 17	62.4	4.0	1843	16	US-10-041-018-5	Sequence 5, Appli
C 18	58.6	3.7	1095	16	US-10-282-122A-40388	Sequence 40388, A
C 19	58.6	3.7	1364	15	US-10-369-493-36669	Sequence 36669, A
C 20	56.4	3.6	6115	15	US-10-311-455-1774	Sequence 1774, Ap
C 21	53.6	3.4	520	17	US-10-021-323-7699	Sequence 7699, Ap
C 22	52.4	3.3	3673778	15	US-10-312-841-2	Sequence 2, Appli
C 23	50	3.2	912	15	US-10-369-493-36606	Sequence 36606, A
C 24	49.6	3.2	961	15	US-10-369-493-36678	Sequence 36678, A
C 25	49.2	3.1	6802	15	US-10-311-455-1204	Sequence 1204, Ap
C 26	48.8	3.1	1044	16	US-10-282-122A-10905	Sequence 10905, A
C 27	48.2	3.1	984	15	US-10-369-493-45179	Sequence 45179, A
C 28	48	3.1	6127	15	US-10-311-455-1587	Sequence 1587, Ap
C 29	47	3.0	383	10	US-09-814-353-18006	Sequence 18006, A
C 30	46.4	3.0	469	17	US-10-021-323-16830	Sequence 16830, A
C 31	46.4	3.0	6681	15	US-10-311-455-128	Sequence 128, App
C 32	46.4	3.0	6681	15	US-10-240-452-4	Sequence 4, Appli
C 33	46.4	3.0	10856	15	US-10-311-455-430	Sequence 430, App
C 34	46.4	3.0	18133	15	US-10-311-455-914	Sequence 914, App
C 35	46.4	3.0	18133	16	US-10-257-166-100	Sequence 100, App
C 36	46	2.9	376	9	US-09-960-352-5087	Sequence 5087, Ap
C 37	45.8	2.9	3673778	15	US-10-312-841-1	Sequence 1, Appli
C 38	45.6	2.9	565	17	US-10-021-323-11125	Sequence 11125, A
C 39	45.4	2.9	1026	15	US-10-369-493-44029	Sequence 44029, A
C 40	45.4	2.9	1029	15	US-10-369-493-24963	Sequence 24963, A
C 41	45.2	2.9	5678	15	US-10-311-455-1111	Sequence 1111, Ap
C 42	45.2	2.9	392000	15	US-10-027-983-11	Sequence 11, Appl
C 43	45.2	2.9	392000	15	US-10-448-753-11	Sequence 11, Appl
C 44	45.2	2.9	465237	9	US-09-933-267A-1	Sequence 1, Appli
C 45	45	2.9	10254	15	US-10-311-455-1048	Sequence 1048, Ap

ALIGNMENTS

RESULT 1

US-10-041-018-1
; Sequence 1, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT APPLICATION NUMBER: US/10/041.018
; PRIOR FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1569
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-041-018-1

Query Match	100.0%	Score 1569;	DB 16;	Length 1569;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1569;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	AATATTACATATAGATATAGACAAAGCCCGCATTTTCATCTGAAAGGTAACCTTCATT	60	
Db	1	AATATTACATATAGATATAGACAAAGCCCGCATTTTCATCTGAAAGGTAACCTTCATT	60	
Qy	61	ATTATAGTGTATCCACGTTTCACGCTTCCAGATAGCAGAAATACGTTGTTTCAT	120	
Db	61	ATTATAGTGTATCCACGTTTCACGCTTCCAGATAGCAGAAATACGTTGTTTCAT	120	
Qy	121	ATGTTATGCTGATCATTTGATGCTTACTACCATTTTCTTGTGCTTCGCTTCCTTTT	180	
Db	121	ATGTTATGCTGATCATTTGATGCTTACTACCATTTTCTTGTGCTTCGCTTCCTTTT	180	

QY 181 GACGTTTTTTGAAGCAAAAAAGTCAAGACAGATGCTTCAAAAACCAATGTAAGCC 240
Db 181 GACGTTTTTTTGAAGCAAAAAAGTCAAGACAGATGCTTCAAAAACCAATGTAAGCC 240
QY 241 TCAATTTTCAAGAAGCTCTAATAGAAAGAGAACAAAGAGTTTACGAGTCTGGAATCA 300
Db 241 TCAATTTTCAAGAAGCTCTAATAGAAAGAGAACAAAGAGTTTACGAGTCTGGAATCA 300
QY 301 ATGGAGGCCAAGATAGATGAGCTGATCAATAATGATCCTGTTGGTCCAGCCAAAATGAA 360
Db 301 ATGGAGGCCAAGATAGATGAGCTGATCAATAATGATCCTGTTGGTCCAGCCAAAATGAA 360
QY 361 AGCTTGATTTCAAAACCTTTATATCACTCCCTTTGAAACCTGGCGAAGAACTTTAGACTA 420
Db 361 AGCTTGATTTCAAAACCTTTATATCACTCCCTTTGAAACCTGGCGAAGAACTTTAGACTA 420
QY 421 AATTTAAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACCAGCTGGCCATAGTT 480
Db 421 AATTTAAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACCAGCTGGCCATAGTT 480
QY 481 TCGCAAAATTTGAGCTCTTGCAATAATCCAGCCTTTTAAATCGACGATATAGAAGATAAT 540
Db 481 TCGCAAAATTTGAGCTCTTGCAATAATCCAGCCTTTTAAATCGACGATATAGAAGATAAT 540
QY 541 GCTCCCTTGAGAGGGGACAGACCCTTCTCACTTAATCTTCGGTGTACCTCCACTATA 600
Db 541 GCTCCCTTGAGAGGGGACAGACCCTTCTCACTTAATCTTCGGTGTACCTCCACTATA 600
QY 601 AACACCGCAAAATTAATGATTTTTCAGAGCCATCAACTTGTATCGCAGCTTAACCAAAA 660
Db 601 AACACCGCAAAATTAATGATTTTTCAGAGCCATCAACTTGTATCGCAGCTTAACCAAAA 660
QY 661 GAGCCTTTGATCATAAATTTGATTTACGATTTTCAAGAGAAATGATCAATCTACATAGG 720
Db 661 GAGCCTTTGATCATAAATTTGATTTTACGATTTTCAAGAGAAATGATCAATCTACATAGG 720
QY 721 GGAAGAGCTTGATATATCTGAGAGACTTTTTCGCTGAAATCATACCTACTCAGGAG 780
Db 721 GGAAGAGCTTGATATATCTGAGAGACTTTTTCGCTGAAATCATACCTACTCAGGAG 780
QY 781 ATGATTTTGAATATGTTTATGAATAAACAAGCGGCTTTTTCAGATTAAGTTGAGACTC 840
Db 781 ATGATTTTGAATATGTTTATGAATAAACAAGCGGCTTTTTCAGATTAAGTTGAGACTC 840
QY 841 ATGGAAGCGCTGTCTCTCTCAACACCGGCAATCGTGGTTCCTTTCAATAATCTT 900
Db 841 ATGGAAGCGCTGTCTCTCTCAACACCGGCAATCGTGGTTCCTTTCAATAATCTT 900
QY 901 CTGGGTATTTATCAGATAGAGATGATTAATGAAATTTGAAAGATTTCCAAATGTC 960
Db 901 CTGGGTATTTATCAGATAGAGATGATTAATGAAATTTGAAAGATTTCCAAATGTC 960
QY 961 AGCGAAAAGGCTTTGCTGAGGACATTTACAGAGGGGAGTTATCTTTCCCATCTGCCAC 1020
Db 961 AGCGAAAAGGCTTTGCTGAGGACATTTACAGAGGGGAGTTATCTTTCCCATCTGCCAC 1020
QY 1021 GCCTTTAACTTCACTAAAAGAAAGTCAAACTGAGCAACAAATGAAATTTCTAAGAAAT 1080
Db 1021 GCCTTTAACTTCACTAAAAGAAAGTCAAACTGAGCAACAAATGAAATTTCTAAGAAAT 1080
QY 1081 CTCTGTTGAGGACAAAGTATAGATATATAAATCAAGCTGATTTCAATCTGGAATTC 1140
Db 1081 CTCTGTTGAGGACAAAGTATAGATATATAAATCAAGCTGATTTCAATCTGGAATTC 1140
QY 1141 GACACCAATTCATTTGGGCTACACCAAAATTTTATTAATCAATTTAGTGAATATGATAAAA 1200
Db 1141 GACACCAATTCATTTGGGCTACACCAAAATTTTATTAATCAATTTAGTGAATATGATAAAA 1200
QY 1201 AATGATATGAAAAATAGTATTTACCTGATTTGGCTTCGATTTCCGACACCGCCACCAAT 1260
Db 1201 AATGATATGAAAAATAGTATTTACCTGATTTGGCTTCGATTTCCGACACCGCCACCAAT 1260

QY 1261 TTACATGCGAATTTGTTATATATATAGACCACTTATCCGAATTTGTGAAATAAATGATC 1320
Db 1261 TTACATGCGAATTTGTTATATATATAGACCACTTATCCGAATTTGTGAAATAAATGATC 1320
QY 1321 AATCAAAATAGTGGAGGAAGATAGTCAGAAATAAGACCTTCTCTCTCTCTCTCTCGCATC 1380
Db 1321 AATCAAAATAGTGGAGGAAGATAGTCAGAAATAAGACCTTCTCTCTCTCTCTCTCGCATC 1380
QY 1381 TATACATACGATTTTCATATATACGTTTCAATTCATCATCTTTTGTATATATCTCAAAAAGA 1440
Db 1381 TATACATACGATTTTCATATATACGTTTCAATTCATCATCTTTTGTATATATCTCAAAAAGA 1440
QY 1441 TCTCTTAGTTTCGCAAAATAGTCAAAATCTTCAAAATTTATATAGCTTTTATATTTTCCACGAT 1500
Db 1441 TCTCTTAGTTTCGCAAAATAGTCAAAATCTTCAAAATTTATATAGCTTTTATATTTTCCACGAT 1500
QY 1501 TTCTGAAACTCTTTTATCATCAGCAGCTTAATGCTAGCGGTTACTGTCAAAATCGCGGTA 1560
Db 1501 TTCTGAAACTCTTTTATCATCAGCAGCTTAATGCTAGCGGTTACTGTCAAAATCGCGGTA 1560
QY 1561 AATTCGCGA 1569
Db 1561 AATTCGCGA 1569

RESULT 2
US-10-369-493-46181
; Sequence 46181, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 46181
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; US-10-369-493-46181

Query Match 64.2%; Score 1008; DB 15; Length 1008;
Best Local Similarity 100.0%; Pred. No. 5.6e-230;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 ATGGAGGCCAAGATAGATGAGCTGATCAATAATGATCCTTTGGTCCAGCCAAAATGAA 360
Db 1 ATGGAGGCCAAGATAGATGAGCTGATCAATAATGATCCTTTGGTCCAGCCAAAATGAA 60
QY 361 AGCTTGATTTCAAAACCTTTATATCACTCTTTTGAACCTGGCAAGAACTTTAGACTA 420
Db 61 AGCTTGATTTCAAAACCTTTATATCACTCTTTTGAACCTGGCAAGAACTTTAGACTA 120
QY 421 AATTTAAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACCAGCTGGCCATAGTT 480
Db 121 AATTTAAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACCAGCTGGCCATAGTT 180
QY 481 TCGCAAAATTTGAGCTCTTGCATTAATTCAGACCTTTTAAATCGACGATATAGAAGATAAT 540
Db 181 TCGCAAAATTTGAGCTCTTGCATTAATTCAGACCTTTTAAATCGACGATATAGAAGATAAT 240
QY 541 GCTCCCTTGAGAGGGGACAGACCCTTCTCACTTAATCTTCGGTGTACCTCCACTATA 600
Db 241 GCTCCCTTGAGAGGGGACAGACCCTTCTCACTTAATCTTCGGTGTACCTCCACTATA 300

601 AACACGGCAAAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 660
Db |
301 AACACGGCAAAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 360
Qy |
661 GAGCCTTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 720
Db |
361 GAGCCTTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 420
Qy |
721 GGACAAAGGCTTGGATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 780
Db |
421 GGACAAAGGCTTGGATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 480
Qy |
781 ATGTAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 840
Db |
481 ATGTAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 540
Qy |
841 ATGTAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 900
Db |
541 ATGTAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 600
Qy |
901 ATGTAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 960
Db |
601 ATGTAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 660
Qy |
661 AGCGAAAAGGCTTGGATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 720
1021 GCCCTTAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 1080
Db |
721 GCCCTTAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 780
Qy |
1081 CTCCTTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 1140
Db |
781 CTCCTTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 840
Qy |
1141 GACACCAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 1200
Db |
841 GACACCAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 900
Qy |
1201 AATGATAATGAAATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 1260
Db |
901 AATGATAATGAAATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 960
Qy |
1261 TTACATGACGAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 1308
Db |
961 TTACATGACGAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 1008

RESULT 3

US-10-450-941-5
; Sequence 5, Application US/10450941
; Publication No. US20040063182A1
; GENERAL INFORMATION:
; APPLICANT: TOKOTA JIOSHKA KABUSHIKI KAISHA
; TITLE OF INVENTION: A METHOD OF PRODUCING PRENYLALCOHOL
; FILE REFERENCE: PH-1444PCT
; CURRENT APPLICATION NUMBER: US/10/450,941
; CURRENT FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: JP2000-401701
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: JP2000-403067
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: JP2001-282978
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 5
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS

; LOCATION: (1)...(1005)
US-10-450-941-5

Query Match

Best Local Similarity 100.0%; DB 16; Length 1008;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 301 ATGGAGGCCAAGATAGATGAGCTGATCAATATGATGCTGTTGGTTCAGAGCCAAATGAA 360
Db |
1 ATGGAGGCCAAGATAGATGAGCTGATCAATATGATGCTGTTGGTTCAGAGCCAAATGAA 60
Qy 361 AGCTTGATTTCAAAACCTTATAATCACATCTCTTTTGAACCTGCGCAAGCACTTTAGACTA 420
Db |
61 AGCTTGATTTCAAAACCTTATAATCACATCTCTTTTGAACCTGCGCAAGCACTTTAGACTA 120
Qy 421 AATTTAATAGTTCAAAATTAACAGAGTTATGAAATTTGCCCAAGACAGCTGGCCATAGTT 480
Db |
121 AATTTAATAGTTCAAAATTAACAGAGTTATGAAATTTGCCCAAGACAGCTGGCCATAGTT 180
Qy 481 TCGCAAAATTTGAGCTCTTCGATTAATTTCCAGCTTTTAAATCGACGATATAGAAATAT 540
Db |
181 TCGCAAAATTTGAGCTCTTCGATTAATTTCCAGCTTTTAAATCGACGATATAGAAATAT 240
Qy 541 GCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCCGGTGTACCTCCACTATA 600
Db |
241 GCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCCGGTGTACCTCCACTATA 300
Qy 601 AACACCGCAAAATTAATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 660
Db |
301 AACACCGCAAAATTAATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 360
Qy 661 GAGCCTTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 720
Db |
361 GAGCCTTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 420
Qy 721 GGACAAAGGCTTGGATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 780
Db |
421 GGACAAAGGCTTGGATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 480
Qy 781 ATGTAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 840
Db |
481 ATGTAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 540
Qy 841 ATGTAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 900
Db |
541 ATGTAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 600
Qy 901 CTGGGTATTTATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 960
Db |
601 CTGGGTATTTATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 660
Qy 961 AGCGAAAAGGCTTGGATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 1020
Db |
661 AGCGAAAAGGCTTGGATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 720
Qy 1021 GCCTTTAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 1080
Db |
721 GCCTTTAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 780
Qy 1081 CTCCTTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 1140
Db |
781 CTCCTTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 840
Qy 1141 GACACCAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 1200
Db |
841 GACACCAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 900
Qy 1201 AATGATAATGAAATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 1260
Db |
901 AATGATAATGAAATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 960
Qy 1261 TTACATGACGAATTTGTAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAAA 1308
Db |

Db 961 TTACATGACGAATCTTTATATATATATAGACCACTTATCCGAATTGTGA 1008

RESULT 4
US-10-189-268-4
; Sequence 4, Application US/10189268
; Publication No. US20040005570A1
; GENERAL INFORMATION:
; APPLICANT: Nicholas M. Dean
; APPLICANT: C. Frank Bennett
; APPLICANT: Kenneth W. Dobie
; TITLE OF INVENTION: ANTISENSE MODULATION OF GERANYLGERANYL DIPHOSPHATE SYNTHASE 1 EXH
; FILE REFERENCE: PTS-0021
; CURRENT APPLICATION NUMBER: US/10/189,268
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 131
; SEQ ID NO 4
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (170)...(1072)
US-10-189-268-4

Query Match 7.3%; Score 114; DB 16; Length 1395;
Best Local Similarity 52.1%; Pred. No. 1.2e-16;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;

QY 359 AAGCTTGGATTTCAAAACCTTATAATACATCCTTTTGAACCTGGCAAGACTTTAGAC 418
Db 195 AAGAAATCTTCTAGAACCTTATAATATCTTACTTACAGTTACAGTTAAACAAGTGAGAA 254

QY 419 TAAATTAATAGTTCAAAATTAACAGAGTTAAGATTTGCCAAAGACCGAGTGGCCATAG 478
Db 255 CCAAACTTTACAGGCAATTAATCATTTGGCTGAAAGTTCCAGAGGACAAAGCTACAGATTA 314

QY 479 TTTCGCAAAATTTGTAGCTCTTCGCAATTTCCAGCCCTTTTAAATCGAGGATATAGAAGATA 538
Db 315 TTATTGAAGTGACAGAAATGTTGCATAATGCCAGTTTACTCATCGATGATTTGAAGACA 374

QY 539 ATGCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTA 598
Db 375 ACTCAAACTCCGACCTGGCTTCCAGTGCCCAAGCAATCATGGAATCCCATCTGTCA 434

QY 599 TAAACACCGCAATTAATGATTTTCAGAGCCATGCAACTTTGTATCGAGCTTAACACAA 658
Db 435 TCAATCTGCCAATTACGTGTATTTCTTGGCTTGGAGAAAGTCT----- 479

QY 659 AAGAGCTTTGTATCATAATTTGATTAAGATTTTCAAGAGAAATTTGATCAATCTACATA 718
Db 480 TAACCTTTGATCACCAGATGCGAGTGAAGCTTTTACCGCCAGCTTTTGGAACTCAATC 539

QY 719 GGGCAAGAGCTTTGGATATATACTGGAGAGACTTTCTGCTGAAATCATACCTACTCAGG 778
Db 540 AGGGAAGAGCTTTAGATATTTTACTGGAGGA-----TAATTACACTTTGCCACTGAAG 593

QY 779 AGATGATTTGAATATGTTTATGAATAAACAAGCGGCTTTTCAGATTTAAAGTTGAGAC 838
Db 594 AAGAATATAAGCTATGCTGTCGAGAAACACAGTGGAGCTTTTGGATTAGCAGTAGGTC 653

QY 839 TCATGGAAGCGCTGCTCTCTCTCTCAACACCGCCATTCGTTGGTTCTTTTCATATAATC 898
Db 654 TCATGAGTTGTTCTCTGATTAACAGAGATTTAAACCGCTACTT-----ATA 704

QY 899 TTCTGGGTATTTATTTATCAGATTTAGATGATTTACTTTGAATTTGAAAGATTTCCAAATG 958
Db 705 CACTTGGGCTCTTTTCCAAATTAGGATGATTTATGCTTAATCTACACTCCAAAGAAATATA 764

QY 959 CAGCGAAAAGGCTTTGCTGAGACATTTACAGAGGGGAGATTTCTTTTCCCATCTGCC 1018
Db 765 GTGAAACAAAGATTTTGTGAAGATCTGACAGAGGGAAGATTTCTATTTCTCTACTATTC 824

QY 1019 ACGCCCTT 1026
Db 825 ATGCTATT 832

RESULT 5
US-09-814-353-21762
; Sequence 21762, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21762
; LENGTH: 1692
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..1592
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-21762

Query Match 7.3%; Score 114; DB 10; Length 1692;
Best Local Similarity 52.1%; Pred. No. 1.3e-16;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;

QY 359 AAGCTTGGATTTCAAAACCTTATAATACATCCTTTTGAACCTGGCAAGACTTTAGAC 418
Db 280 AAGAAGATTTCTTAGAACCTTATAATCTTACTTCAGTTTACCAGGTAAACAAGTGAGAA 339

QY 419 TAAATTTAATAGTTTCAAAATTAACAGAGTTATGAATTTGCCAAAGACCCAGCTGGCCATAG 478
Db 340 CCAAACTTTCAGAGGCAATTTATCATTTGGCTGAAAGTTCCAGAGGACAGCTACAGATTA 399

QY 479 TTTCGCAAAATTTGTAGCTCTTCGCAATTTCCAGCCCTTTTAAATCGACGATATAGAAGATA 538
Db 400 TTATTGAAGTGACAGAAATGTTGCATAATGCCAGTTTACTTCATCGATGATTTGAAGACA 459

QY 539 ATGCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCCGTTGTTACCTCCACTA 598
Db 460 ACTCAAACTCCGACCTGGCTTTCCAGTGGCCCAACAGCATCTATGGAATCCCATCTGTCA 519

QY 599 TAAACACCGCAAAATTTATGATTTTTCAGAGCCATGCAACTTTGTATCGAGCTAACCAAA 658
Db 520 TCAATTTCTGCAATTTACGTTATTTCTTGGCTTGGAGAAAGTCT----- 564

QY 659 AAGAGCTTTGTATCATAATTTGATTTACGATTTTTCAGAGAAATTCATCAATCTACATA 718
Db 565 TAAACCTTGTATCACCAGATGCGAGTGAAGCTTTTACCCGCCAGCTTTTGGAACTCCATC 624

QY 719 GGGGACAGAGCTTTGGATATATATCTGGAGAGACTTTTCTGCTGAAATCATACCTACTCAGG 778

Db 625 AGGACAGAGCCCTAGATATTTACTGGAGGA-----TAATTACACTTGTCCCACTGAAG 678
QY 779 AGATGTAATTTGAATATGGTTATGAATAAAACAGGCGGCTTTTCAGATTAAACGTTGAGAC 838
Db 679 AAGATATAAAGCTATGGTCTGCAGAAAACAGGTGACTGTTTGGATTAGCAGTAGGTC 738
QY 839 TCATGAAGCGGTGTCTCCTCTCCTCACACCGGCCCAATTCGTTGGTTCTTTTCATAAAATC 898
Db 739 TCATGCAAGTTGTTCTCTGATTAACAAGAGATTTAAACCGCTACTT-----AATA 789
QY 899 TTCTGGGTATTTATCAGATTAGAGATGATTAATCTTGAATTTGAAAGATTTCCTCAAAATG 958
Db 790 CACTTGGGCTCTTTTCCAAATTAGGATGATTAATCTTAACTCACTCCAAAGATATA 849
QY 959 CCAGCGAAAAGGCTTTGCTGAGGACATTAACAGAGGGGAAGTTATCTTTCCCATCGTCC 1018
Db 850 GTGAAAACAAAAGTTTCTGAGATCTGACAGAGGGAAGTTCTCAATTTCTACTATTC 909
QY 1019 AGGCCCTT 1026
Db 910 ATGCTATT 917

RESULT 6
US-10-041-018-3
; Sequence 3, Application US/10041018
; Publication No. US2004007232A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; PRIOR FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 2534
; TYPE: DNA
; ORGANISM: Human
US-10-041-018-3

Query Match 7.3%; Score 114; DB 16; Length 2534;
Best Local Similarity 52.1%; Pred. No. 1,6e-16;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;

QY 359 AAAGCTTGATTTCAAACCTTTATATACATCCCTTTTGAAACCTGGCAAGAACTTTAGAC 418
Db 258 AAAGAAATTTCTTAGAACCTTATAAATCTTACTTACAGTTACCAAGTAAACAAGTGAGAA 317
QY 419 TAAATTTAAATAGTTCAAATTAACAGATTAATGAAATTTGCCCAAGACCAAGTGGCCATAG 478
Db 318 CCAAACTTTCACAGGATTTAATCATGCTGGAAGTTCCAGAGGACAAAGTACAGATTA 377
QY 479 TTTCGCAATTTGTTAGCTCTTGATTAATTCAGAGCTTTTAAATCGACATATAGAATA 538
Db 378 TTATTGAAGTGACAGAAATGTCATATATGCGATTTACTCATCGATGATATGAAGACA 437
QY 539 ATGCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCCCTCCACTA 598
Db 438 ACTCAAACTCCGACGTGGCTTTCCAGTGGCCACAGCATCTATGGAATCCCATCTGCA 497
QY 599 TAAACCGCAAAATATATGATTTTCAGAGCCATGCACTTGATTCGAGCTTAACCAACAA 658
Db 498 TCAATTCGCAATATACGTGATTTCTTGGCTTGGAGAAAGTCT-----542
QY 659 AAGAGCCCTTTGATCATAAATTTGATTTACGATTTTCAACGAGAAATTCATCAATCTACATA 718
Db 543 TAAACCTTGATCACCAGATGCGAGTGTATTTTACCGGCGAGCTTTTGGAACTCCATC 602
QY 719 GGGACAAAGCTTGGATATATACCTGGAGAGACTTTTCTGCTGAAATCATACCTACTCAGG 778

RESULT 7

US-09-925-302-171
; Sequence 171, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PAL04
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 171
; LENGTH: 2529
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-302-171

Query Match 7.2%; Score 113.6; DB 9; Length 2529;
Best Local Similarity 51.9%; Pred. No. 2e-16;
Matches 347; Conservative 1; Mismatches 290; Indels 30; Gaps 3;

QY 359 AAAGCTTGATTTCAAACCTTTATATACATCCCTTTTGAAACCTGGCAAGAACTTTAGAC 418
Db 246 AAAGAAATTTCTTAGAACCTTATAAATCTTACTTCAAGTTACCAAGTAAACAAGTGAGAA 305
QY 419 TAAATTTAAATAGTTCAAATTAACAGATTAATGAAATTTGCCCAAGACCAAGTGGCCATAG 478
Db 306 CCAAACTTTCACAGGCAATTTAATCATTTGGCTGAAAGTTCCAGAGGACAAAGTACAGATTA 365
QY 479 TTTCGCAATTTGTTAGCTCTTGATTAATTCAGAGCTTTTAAATCGACATATAGAATA 538
Db 366 TTATTGAAGTGACAGAAATGTCATATATGCGAGTTTACTCATCGATGATATGAAGACA 425
QY 539 ATGCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCCCTCCACTA 598
Db 426 ACTCAAACTCCGACGTGGCTTTTCCAGTGGCCACAGCACTATGGAATCCCATCTGCA 485
QY 599 TAAACCGCAAAATATATGATTTTCAGAGCCATGCACTTGATTCGAGCTTAACCAAA 658
Db 486 TCAATTCGCAATATACGTGATTTCTTGGCTTGGAGAAAGTCT-----530
QY 659 AAGAGCCCTTTGATCATAAATTTGATTTACGATTTTCAACGAGAAATTCATCAATCTACATA 718
Db 531 TAAACCTTGATCACCAGATGCGAGCTTTTACCCGCCAGCTTTTGGAACTCCATC 590

Qy	719	GGGACAAAGCTTGGATATATACATGGAGACATTTCTGCCTGAAAATCATACCTACTCAGG	778
Db	591	AGGACAAGGCTAGATATTTACTGGAGGGA-----TAATTAACATGTGTCCTCCACTGAAG	644
Qy	779	AGATGTAATTTGAATATATGGTTATGAAATAAACACAGCGGCCCTTTTCAGATTAACTGTGAGAC	838
Db	645	AAGAATAATAAAGCTATGTGCTGCAGAAAAACAGGTGGACTGTTTGGATTAGCAGTAGGTC	704
Qy	839	TCATGGAGAGCGCTCTCTCCCTCCATCACACACGGCCCATTCGTGTGGTTTCCTTTCATATAATC	898
Db	705	TCATGCGAGTGTCTTCTCTGATTTACAAAGAAGTTTAAACCGCTACTTT-----AATA	755
Qy	899	TTCTGGGTATTTATTTATTCAGATTAGATGATTAATCTTGAATTTGAAAGATTTTCCAAATGT	958
Db	756	CAC TTGGGCTCTTTTCCAAATTAGGATGATTAATGCTAATCTACACTCCAAAGATATA	815
Qy	959	CCACGCAAAAAGGCTTGTCTGAGGACATTACAGAGGGGAAGTATCTTTTCCCATCTCTCC	1018
Db	816	GTGAAAAACAAAGTTTCTGTGAAGATCTGCAGAGGGGAAGTTCTCATTTCTCTACTATTC	875
Qy	1019	ACGCCCTT	1026
Db	876	ATGCTATT	883

```

RESULT 8
US-09-925-302-171
; Sequence 171, Application US/09925302
; Publication No. US20030064072A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 171
; LENGTH: 2529
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-302-171

```

D _b	531	TACCCCTTGATCACCAGATGCGAGTGAAGCTTTTTCACGCCGCAGCTTTTGGAACTCOATC	590
Q _y	719	GGGGACAAGGCTTGGATATATACTCGAGAGACATTTCTGCTGAAATCATACCTACTCACGG	778
D _b	591	AGGGACAAGGCCGTAGATTATTTCTCGGAGGA-----TAATTACACTTGTCTCCCACTAAG	644
Q _y	779	AGATGTATTTGAATATGGTTTATGAATAAAAACAGCGGCCCTTTTCAGATTHAACGTTGAGAC	838
D _b	645	AAGAATATAAAGCTATGGTCTGCAGAAAAACAGGTGCACTGTTTGGATTACAGTAGTGGTC	704
Q _y	839	TCATGGAAGCGCTGCTCCTTCTCTCACACAGGCCCAATTCGTTGGTTCTTTTCATATAATC	898
D _b	705	TCATGCGAGTTGTTCTCTGATTAACAAGAGATTTAAACCGCTACTT-----AATA	755
Q _y	899	TTCTGGGTATTTATTTACAGATTAGACATGATTACTTTGAATTTTGAAGATTTCCAATGT	958
D _b	756	CAC TTGGGCTCTTTTTCCAAATTAGGATGATTATGCTAATCTACACTCCAAAGAAATATA	815
Q _y	959	CCACGGAAAAGGCTTTGCTGAGACATTTACAGAGGGGAAAGTTATCTTTTCCCACCTGCC	1018
D _b	816	GTGAAAACAAAGTTTTGTGTGAAGATCTGCAGAGGGGAAAGTTCTCATTTCTTACTATTTC	875
Q _y	1019	ACGCCCTT	1026
D _b	876	ATGCTATT	883

```

RESULT 9
US-10-369-493-36092
; Sequence 36092, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 36092
; LENGTH: 1044
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-10-369-493-36092

```

Query Match	6.5%;	Score 102.6;	DB 15;	Length 1044;
Best Local Similarity	49.9%;	Mismat. No. 5.5e-14;		
Matches 363;	Conservative	0;	Mismatches 334;	Indels 30; Gaps 3
300	AATGAGGCCAAGATAGATGAGCTCATCAATAATGATCTGTTCGTTCAGGCCAAAATCA	359		
153	AAATCAGTCAAGTCTCGACGGGACAAATACAAAGATGGGAAATGGTCCCAAGAGAAATCA	212		
360	AAGCTTGATTTCAAACCTTATTAATACATCTCTTTTGAAACCTGCGACAGACTTTTAGACT	419		
213	GGAGGTGATCATGGGTCCGTAGCATGTACATGTGCAACACCCGGGGAAGGACTCGACG	272		
420	AAATTTAATAGTTCAAATTTAAACAGAGTTATGAATTTGCCCAAAGACGAGCTGGCCATAGT	479		
273	GCAGATGATCAACGGCTTTTAAAGTATGGTTGAAGGTGCCATCTGAGAGCGCTGGCCATCAT	332		
480	TTTCGCAAAATGTTGAGCTCTTCGCATAAATTCACAGCCCTTTTAATTCGACGATATAGAAGATAA	539		
333	CACCAAGTAGTGCTTATGCTCCATACCGCTTCATTATTGATCGACGAGCTCGAAGACAA	392		
540	TGCTCCCTTGAGAGGGGACAGACCACCTTCTCACITTAATCTTCGGTGTACCTCCACTAT	599		


```

Db 15443 GACAACTCAAACCTCGAGTGGCTTTCCAGTGGCCACAGCATCTATGGAATCCACT 15502
Qy 595 ACTATAACACCGCAAAATATATATATTTTACAGAGCCATGCACTTGTATCGAGCTAAC 654
Db 15503 GTACATATCTCGCAATATACGTATATTTCTTGGCTTGGAGAAAGTCTTAACCCCT- 15558
Qy 655 ACAAGAAGCCCTTTGTATCATATTTGATATGATTTTCAAGAAAGATTTGATCAATCTA 714
Db 15559 -----TGATACCCAGATGAGTGAAGCTTTTACCCGCCAGCTTTTGAACATC 15607
Qy 715 CATAGGGAACAGGCTTGATATATATGAGAGACTTTCTGCTGAAATCATACCTACT 774
Db 15608 CATCAGGACAGGCTAGATATTTACTGAGGATATATACATG-----TCCACT 15661
Qy 775 CAGGAGATGATTTGATATGATTTGATATGATTTGATATGATTTGATATGATTTGAT 834
Db 15662 GAAGAAGATATAAAGCTATGCTGCTGCAAGAAACAGGTGGAGTGTGATTTAGCAGTA 15721
Qy 835 AGACTCATGAGAGCGCTGCTCTCTCTCTCACACACCGGCCATTTGGTTGCTCTTTCATA 894
Db 15722 GGTCTCAGCAGTGTGTTCTCTGATTAAGAGAG-----TTTAAACCGCTACTT 15772
Qy 895 AATCTTCTGGTATATTTATCAGATTAGAGATGATTTACTTGAATTTGAAAGATTTCCAA 954
Db 15773 AATACACTTGGCTCTTTTCCAAATAGGATGATTTATGCTAATCTACACTCCAAAGAA 15832
Qy 955 ATGTCAGCGAAAGAGCTTTGCTGAGGACATTTACAGAGGAGGATTTCTTTCCCATC 1014
Db 15833 TATAGTGAAGAAACAAAGTTTTGTGAAGATCTGACAGAGGAAAGTTCTCATTTCTACT 15892
Qy 1015 GTCCAGCGCCCTT 1026
Db 15893 ATTCATGCTATT 15904

RESULT 12
US-10-041-018-21
; Sequence 21, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080U1/10025547
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 2222
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-10-041-018-21

Query Match 5.4%; Score 84.2; DB 16; Length 2222;
Best Local Similarity 50.5%; Pred. No. 1.9e-09;
Matches 330; Conservative 0; Mismatches 293; Indels 30; Gaps 4;

Qy 374 AACCTTATATACATCTCTTTTGAACCTGGCAAGAACTTTAGACTAAATTAATAGTTC 433
Db 372 AGCCCTTTACATACATACACAGAGTCTTGGCAAGCAATTCGGCTCTGAGTTGGCTTGG 431
Qy 434 AAATTAACAGAGTTATGAAATTTGCCAAAGCCAGCTGGCCATAGTTTCGGAATTTGTTG 493
Db 432 CTTTCAATCACTGGTTGCTCATACCGGGCGGAAAGTTGGCGCAGATCGGAGACATTTGTC 491
Qy 494 AGCTCTTCATATTCAGCCCTTTTATCAGAGATATAGAGATAATGCTCCCTTGAGAA 553
Db 492 AGATGCTGCACAAATTCAGTTTGTCTCATTTGATGATTTGAGAGCAATTCGATCTCCGCA 551
Qy 554 GGGGACAGACCACTTCTCACTTAATCTTTCGGTGTACCTCCACTATAAACACCGCAAT 613

```

```

Db 552 GAGGTGTCCCGTGGCGCAATTCATCTACGGCGTGGCCAGCACCAATAAATGGGCAACT 611
Qy 614 ATATCTATTTACAGAGCCATGAACTTTGTATCGAGCTAACACACAAAAGAGCTTTGTATC 673
Db 612 ATGCACTCTTTCTGGCGCTGGAGAAGGTGACAGCTGGATCATCCGAGGCT----- 664
Qy 674 ATAAATTTGATTTACGATTTTCAACGAAAGATTTGATCAATCTACATAGGGGACAGGCTTGG 733
Db 665 -----ACCAAGGTGTACACCGCAATTTGCTGGAGCTGCACCGTGGACAGGGCATGG 716
Qy 734 ATATATATCTGGAGAGACTTTTCTGCTGCTGAAATCTATACCTACTCTCAGGAGATGATTTGAATA 793
Db 717 AGATCTATTTGGCGCACAGCTT-----CACGTGTCCATCCGAGTCCGATTAACAAGTGA 770
Qy 794 TGGTTATGAATAAAACAGCGGCCCTTTTTCAGATTAAACGTTTGAGACTCATGGAAGCGCTGT 853
Db 771 TGACTGTGGGCAAACTGGCGCCCTCTTTATGCTGGCCATTCGCTTATGCA--GCTGT 827
Qy 854 CTCCTTCTCTCACACACCGGCCATTCGTTGGTTTCTTTTCAATAATCTTCTGGGTATTTT 913
Db 828 TCAGCTTCCAAACAGGAGGACTTATCGAAG-----TTGACGGCTATATTTGGGCTGTACT 881
Qy 914 ATCAGATTAGAGATGATTTGATTTGAATTTGAAGATTTCCAAATGTCCAGCAAAAGGCT 973
Db 882 TTCAGATACGCGACGACTATTGCAATCTGAGTCTGAAAGAGTACACGGAGAACAGAGCT 941
Qy 974 TTGCTGAGCACATTACAGAGGGGAGATTTATCTTTTCCCATCGTCCACGCGCCTT 1026
Db 942 TGCCGAGGACTTGACGAGGAGGAGTTCGGCTTCCGGTAAATCCATGCGGTT 994

RESULT 13
US-10-001-192A-14
; Sequence 14, Application US/10001192A
; Publication No. US20040091958A1
; GENERAL INFORMATION:
; APPLICANT: Oeljen, Albert
; APPLICANT: Verdoes, Jan
; APPLICANT: Wery, Jan
; TITLE OF INVENTION: IMPROVED METHODS FOR TRANSFORMING
; TITLE OF INVENTION: PHAFFIA STRAINS, TRANSFORMED PHAFFIA STRAINS SO OBTAINED AND
; FILE REFERENCE: 24615-20104.01
; CURRENT FILING DATE: 2002-06-11
; PRIOR APPLICATION NUMBER: US/10/001,192A
; PRIOR FILING DATE: 1995-12-22
; PRIOR APPLICATION NUMBER: EP95203620.0
; PRIOR FILING DATE: 1996-04-11
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1882
; TYPE: DNA
; ORGANISM: Phaffia rhodozyma
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (82)...(1212)
US-10-001-192A-14

Query Match 4.5%; Score 70.6; DB 16; Length 1882;
Best Local Similarity 51.8%; Pred. No. 3.1e-06;
Matches 160; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

Qy 352 CAAATGAACCTTGATTTCAAACCTTATAATCACATCTTTTGAACCTTGCAAGAAC 411
Db 130 CAGATGATATCGTCTCTCTGTAACCTGTATCATCTACCTAGAAAGAACCTTGGAAGAA 189
Qy 412 TTTAGACTAAATTTAATAGTTTCAATTAACAGAGTTATGATTTGCCAAAGACAGCTG 471
Db 190 ATTGATCACAACTCATCGAGGCTTTCAACTATTTGGTTGGATGTCAAGAGAGGATCTC 249

```

QY 472 GCATAGTTTCGCAAAATGTTAGCTCTTGCATAAATCCAGCCCTTTTAAATCGACGATATA 531
Db 250 GAGGTCTATCCAGAACGTTGTTGCATGCTACATACCGCTAGCTTATTAATGACGATGTG 309
QY 532 GAAGATAATGCTCCCTTGAGAGGGGACAGACCACTTCTCAGCTTAATCTTCGGTGTACCC 591
Db 310 GAGGATTCATCGGTCTCTCAGGCGTGGTGCCTGTGCGCCATCTAATTAACGGGATTCGG 369
QY 592 TCCACTATATAACACCGCAAAATATATGTTATTTCCAGAGCCATGCAACTTGTATCGCAGCTA 651
Db 370 CAGACATATAACACTGCAAACTACGCTCTACTTTCTGGCTTATCAGAGATCTTCAAGCTT 429
QY 652 ACACAAAA 660
Db 430 CGCCCAACA 438

RESULT 14
US-09-969-034-1324/c
; Sequence 1324, Application US/09969034
; Publication No. US20040110668A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Astle, Jon H.
; APPLICANT: Carroll, Eddie III
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dwivedi, Poornima
; APPLICANT: Molino, Gary A.
; APPLICANT: Thiagalingam, Arunthathi
; APPLICANT: Lewis, Marcia E.
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially
; FILE REFERENCE: 1657/1032
; CURRENT APPLICATION NUMBER: US/09/969,034
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/237,271
; NUMBER OF SEQ ID NOS: 4494
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1324
; LENGTH: 554
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 496..517, 531, 536
; OTHER INFORMATION: n = A,T,C or G
US-09-969-034-1324

Query Match 4.4%; Score 68.8; DB 11; Length 554;
Best Local Similarity 54.2%; Pred. No. 4.6e-06;
Matches 192; Conservative 0; Mismatches 147; Indels 15; Gaps 2;
QY 673 CATTAATTGATTACGATTTTCAACGAGAAATGATCAATACATACATAGGAGCAAGGCTTG 732
Db 422 CAAGATGCAAGTGAAGCTTTTACCGCGCAGCTTTTGGAACTCCATCAGGCAAGGCGCTA 363
QY 733 GATATATCTGGAGAGACTTTCGCTGCAATCATACCTACTCAGGAGATGTAATTGAAT 792
Db 362 GATATTACTGGAGGA-----TAATTACACTTGTCCCACTGAAGAAGATATAAAGCT 309
QY 793 ATGTTTATGAATAAACAAGCGGGCTTTTCAGATTAAAGTTGAGACTCATGGAAGCGGTG 852
Db 308 ATGGTGTGCGAATAACAGGTGGACTGTTTGGATTAGCAGTAGGTCTCATGCTGTTTC 249
QY 853 TCTCTTCTCCTCACACCGGCGCATTCGTTGGTTCCTTTCATAAATCTTCTGGGTATTATT 912
Db 248 TCTGATTACAAAGAAGATTAAACCGCTACTT-----AATACACTTGGGCTCTTT 198
QY 913 TATCAGATTAGAGATGATTACTTGAATTTGAAGATTTCCAAATGTCCAGCGAAAGGCG 972
Db 197 TTCCAAATTAGGATGATTGCTAATCTACACTCCAAAGAAATGATAGTGAAACAAAGT 138

QY 973 TTTGCTGAGACATACAGAGGGAAGTTATCTTTTCCCATCGTCCACGCCCTT 1026
Db 137 TTTTGTGAAGATGTGACAGAGGGGAAAGTTCTCATTTCTACTATTATGCTATT 84
RESULT 15
US-09-969-034-1924/c
; Sequence 1924, Application US/09969034
; Publication No. US20040110668A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Astle, Jon H.
; APPLICANT: Carroll, Eddie III
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dwivedi, Poornima
; APPLICANT: Molino, Gary A.
; APPLICANT: Thiagalingam, Arunthathi
; APPLICANT: Lewis, Marcia E.
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially
; FILE REFERENCE: 1657/1032
; CURRENT APPLICATION NUMBER: US/09/969,034
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/237,271
; NUMBER OF SEQ ID NOS: 4494
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1924
; LENGTH: 716
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 519, 560, 581, 585, 590, 600, 602, 628, 669, 670, 673, 679,
; LOCATION: 702
; OTHER INFORMATION: n = A,T,C or G
US-09-969-034-1924

Query Match 4.4%; Score 68.4; DB 11; Length 716;
Best Local Similarity 54.7%; Pred. No. 6.5e-06;
Matches 188; Conservative 0; Mismatches 141; Indels 15; Gaps 2;
QY 683 TTACGATTTTCAACGAGAAATGATCAATCTACATAGGGGACAGGCTTGGATATATACT 742
Db 409 TGAAGCTTTTACCGCGCAGCTTTTGGAACTCCATCAGGGAAGGCGCTAGATATTTACT 350
QY 743 GGAGAGACTTTTCGCTGAAATCATACCTACTCAGGAGATCTATTGTAATATGTTATGA 802
Db 349 GGAGGA-----TAATTACACTTGTCCACTGAGAAGATATAAGCTATGGTGCTGC 296
QY 803 ATAAAAAGCGCGGCTTTTCAGATTAAAGTTGAGACTCATGGAAGCGCTGTCTCTCTTCT 862
Db 295 AGAAAAACAGGTGGACTGTTTGGATTAGCAGTAGGTCTCATGAGTTGTTCTCTGATTACA 236
QY 863 CACACCCAGCGCCATTCGTTGGTTCCTTTCATAAATCTCTGGGTATTATTTATCAGATTA 922
Db 235 AAGAAGATTAAACACCGCTACTT-----AATACACTTGGGCTCTTTTCCAAATTA 185
QY 923 GAGATGATTACTTGAATTTGAAGATTTCCTCAAAATGTCCAGCGAAAAAGGCTTGTCTGAGG 982
Db 184 GGGATGATTATGCTAACTACACTCCAAAGATATAGTGAAGAAACAAAGTTTGTGAAG 125
QY 983 ACATTACAGAGGGGAGTTATCTTTTCCCATCGTCCACGCCCTT 1026
Db 124 ATCTGACAGAGGGGAAAGTTCCTCAATTCCTACTATTATGCTATT 81

Search completed: October 17, 2004, 08:40:44
Job time : 735.274 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 17:27:39 ; Search time 4877.9 Seconds
(without alignments)
11721.026 Million cell updates/sec

Title: US-10-041-018-1
Perfect score: 1569
Sequence: 1 aatattacatatagatag.....aatcggttaattcgga 1569

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	519.6	33.1	957	9	CNS06GUS	AL398163 T7 end of
2	510.8	32.6	990	9	CNS06GEP	AL397607 T3 end of
3	468.2	29.8	1021	9	CNS06GIG	AL397742 T3 end of
4	411.8	26.2	537	8	AQ502594	AQ502594 V52B7 mtn
5	195	12.4	1115	9	CNS06CKP	AL392639 T7 end of
6	114	7.3	903	9	AY407652	AY407652 Homo sapi
7	114	7.3	1076	5	EX460174	EX460174 BX460174
8	114	7.3	1417	3	CR612465	CR612465 full-leng
9	114	7.3	2703	3	CR603114	CR603114 full-leng
10	113.6	7.2	903	9	AY407654	AY407654 Mus muscu
11	113.6	7.2	1341	3	AK011980	AK011980 Mus muscu
12	103	6.6	815	4	BI520491	BI520491 603071690
13	99	6.3	905	5	EX439729	EX439729 BX439729
14	97.6	6.2	871	4	BI560099	BI560099 603253287
15	96.2	6.1	974	4	BG675075	BG675075 602621320
16	95.2	6.1	748	2	BF610750	BF610750 dd5h05.y
17	95.2	6.1	1116	3	AY222995	AY222995 Schistos
18	94.8	6.0	1077	4	BG114452	BG114452 602285755
19	94	6.0	514	1	AA573392	AA573392 nm53b07.s
20	94	6.0	610	7	CR360928	CR360928 170005325
21	93.6	6.0	544	2	BF400767	BF400767 UI-R-CAO-
22	93.6	6.0	546	2	BE113864	BE113864 UI-R-CAO-
23	93.6	6.0	678	6	CF146873	CF146873 UI-HF-CB0
24	93.6	6.0	693	4	BM721020	BM721020 UI-E-B00-

25	93.6	6.0	812	5	BX846096	BX846096 BX846096
26	92.8	5.9	803	9	AY407653	AY407653 Pan trogl
27	92.8	5.9	873	6	CD108157	CD108157 AGENCOURT
28	92.8	5.9	911	5	BQ433460	BQ433460 AGENCOURT
29	92.2	5.9	698	7	CN360927	CN360927 170006000
30	91.2	5.8	924	5	BX358978	BX358978 BX358978
31	91	5.8	629	6	CB272278	CB272278 mai56e01.
32	90.4	5.8	569	2	AW270970	AW270970 xs07a01.x
33	89.6	5.7	797	5	BU568661	BU568661 AGENCOURT
34	89.4	5.7	1007	5	BQ886489	BQ886489 AGENCOURT
35	88.8	5.7	515	6	CA880328	CA880328 K0982B05-
36	88.8	5.7	539	6	CA870528	CA870528 K0902G03-
37	88.8	5.7	572	6	CF198250	CF198250 maj46a06.
38	88.8	5.7	777	6	CB724192	CB724192 UI-M-FY0-
39	88.2	5.6	529	9	CG619892	CG619892 OST316302
40	88.2	5.6	807	7	CK476319	CK476319 AGENCOURT
41	88	5.6	630	6	CB190752	CB190752 pl30e10.y
42	87.8	5.6	907	5	BX777578	BX777578 BX777578
43	87.8	5.6	527	6	CB296588	CB296588 12B22037
44	87.8	5.6	859	5	BQ938817	BQ938817 AGENCOURT
45	87.2	5.6	627	6	CB215219	CB215219 NISC_np03

ALIGNMENTS

CNS06GUS 957 bp DNA linear GSS 30-NOV-2001
T7 end of clone AS0AA007D07 of library AS0AA from strain CLIB 533
of Saccharomyces bayanus, genomic survey sequence.

AL398163.1 GI:12151423

GSS.

Saccharomyces bayanus

Saccharomyces bayanus

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

1 (bases 1 to 957)

Souciat, J.L., Aigle, M., Artiguenave, F., Blandin, G.,

Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,

de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,

Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S.,

Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,

Wincker, P. and Weissenbach, J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of

yeast species for molecular evolution studies

FEBS Lett. 487 (1), 3-12 (2000)

20584711

11152876

2 (bases 1 to 957)

Bon, E., Neuvéglise, C., Casaregola, S., Artiguenave, F., Wincker, P.,

Aigle, M. and Durrens, P.

Genomic exploration of the hemiascomycetous yeasts: 5.

Saccharomyces bayanus var. uvarum

FEBS Lett. 487 (1), 37-41 (2000)

20584715

11152880

3 (bases 1 to 957)

Genoscope.

Direct Submission

Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,

2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :

seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This GSS is part of a random genomic sequencing program of thirteen

yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces

exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,

Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces

lactis, Kluyveromyces marxianus var. marxianus, Pichia

angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,

Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to

5 kb were prepared and both extremities were sequenced. See

keywords for description of this sequence and for the sequence of

QY	1350	ATAAAGCCTTCTCTCCTCTCTCTCGCATCTATACATACAGATTTCATATATACGTTTCA	140
Db	684	AAAAGCAGACCTCTCTCCCTCTTTTGCAITTCATATATACAGTATTTATTATACGATTA	743
QY	1410	TTGCATCATCTTTTGATATATATCTCAAAAAGATCTCTTAGTTCGCAATAGTCAAAATCTC	1469
Db	744	TTGCATCATCTTTTGATATATTTCAATAAATCTTTTAGTTGCGATATAGTCAAAATCTC	803
QY	1470	AAATTTATAGCCTTTATATATTTTTCACGATTTCTGAAACTCTCTTTTATCAGCACCGTT	1529
Db	804	AAATTTGCAGCCTTTGATTTTCTTCCATGACTTTTGGAITTCCTTTCTTCAATACCGCT	863
QY	1530	AATGC-----TAGCGTTACTGTCAAAATCGCGGTAATAATTCGCGA	1569
Db	864	AACACTTGCATTTGGCGCTACTGTCAAGTCACTGTAAATTCAGA	909

RESULT 2	CNS06GEP	990 bp	DNA	linear	GSS 30-NOV-2001
LOCUS	T3 end of clone AS0AA004A12 of library AS0AA from strain CLIB 533				
DEFINITION	Of Saccharomyces bayanus, genomic survey sequence.				
ACCESSION	AL397607				
KEYWORDS	AL397607.1 GI:12150371				
SOURCE	GSS.				
ORGANISM	Saccharomyces bayanus				
REFERENCE	Saccharomyces bayanus				
AUTHORS	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
	1 (bases 1 to 990)				
	Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Liorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaita,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissbach,J.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies				
JOURNAL	YEAST				
MEDLINE	FEMS Lett. 487 (1), 3-12 (2000)				
PUBMED	20584711				
REFERENCE	11152876				
AUTHORS	2 (bases 1 to 990)				
	Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P., Aigle,M. and Durrens,P.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 5.				
JOURNAL	Saccharomyces bayanus var. uvarum				
MEDLINE	FEMS Lett. 487 (1), 37-41 (2000)				
PUBMED	20584715				
REFERENCE	11152880				
AUTHORS	3 (bases 1 to 990)				
TITLE	Genoscope.				
JOURNAL	Direct Submission				
	Submitted (07-SEP-2000) Genoscope - Centre National de Sequençage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.				
FEATURES	Location/Qualifiers				
source	1..990				
	/organism="Saccharomyces bayanus"				
	/mol_type="genomic DNA"				
	/strain="CLIB 533"				
	/variety="uvarum"				
	/db_xref="taxon:4931"				

```

/clone_lib="AS0AA04A12"
/clone="end : T3"
<36. .>670
/note="similar to Saccharomyces cerevisiae ORF YPL069c [
BTS1 ; geranylgeranyl diphosphate synthase ]
1 putative frameshift(s)
/evidence=not_experimental
complement(<741. .>989)
/note="similar to Saccharomyces cerevisiae ORF YPL070w [
weak similarity to Vps9p ]"
/evidence=not_experimental

misc_feature
ORIGIN
Query Match 32.6%; Score 510.8; DB 9; Length 990;
Best Local Similarity 75.8%; Pred. No. 3.9e-112;
Matches 717; Conservative 0; Mismatches 182; Indels 47; Gaps 5;

QY 631 ATGCAACTGTGATCGACGCTACACAAAGAGCCTTTGTTATCATATTAATTCAGATT 690
Db 1 ATGCACTCGTATCGCAACTAATGTC-AGACCGTGTCTTACCGTAAATTAATGACATC 59

QY 691 TTCAACGAAGAATTGATCAATCTCATAGGGGACAAGCTTTGGATATATCTCGAGGAC 750
Db 60 TTTAATGAAGAAGTGAATTAATCTACATAGAGCCAGGTTGGATATATATTTGAGAGAT 119

QY 751 TTTCTGCTGAAATCATACCTACTCAGAGAGATGATTGTAATATGTTTGAATATAAACA 810
Db 120 TTTCTACTGAAATCATACCTACTCAGAGAGATGATTGTAATATGTTTGAATATAAACA 870

QY 811 GCGCGCTTTTCAGATTAACGTCAGACTCATGGAAGCGTGTCTCTCTCTCACCAC 870
Db 180 GTTGGTCTTTTCAGATTAACGTCAGACTCATGGAAGCGTGTCTCTCTCTCACCAC 239

QY 871 GGCATTCGTTGGTCTCTTATATAAATCTTCTGGGTATTTATCATAGATGATGAT 930
Db 240 GTTCATTCGCTGGTTCATTCATATAATCTTTAGTATGATTTATCATAGATGATGAT 299

QY 931 TACTTGAATTTGAAGATTTCCAAATGTCGAGGAAAGGCTTTGCTGAGGACATTACA 990
Db 300 TATCTAACTGAAAGACCTTCAAGATGTCCTCAATGAAAGGCTTCGCTGAAGATATCACC 359

QY 991 GAGGGGAAGTATCTTTTCCCATCTGTCACGCGCTTAACTTCATCAAAAGGTCAC 1050
Db 360 GA-AGGAAGCTCTTTCCCATCTGTCATGCTCTTAAATTTCCACGAGGCGCAACGTC 418

QY 1051 ACTGAGCAACAAATGAAATCTTCTGTTGAGGCAAGTGTATAAGATATA 1110
Db 419 ACTGAGCAACAAATGAAATCTTCTGTTGAGGCAAGTGTATAAGATATA 478

QY 1111 AAATAAGCTGATTCAAATCTGGAATTCGACCAATTCATTCGCTTACACCAAAAT 1170
Db 479 AAATAAGCTGATTCAAATCTGGAATTCGACCAATTCATTCGCTTACACCAAAAT 538

QY 1171 TTTTATTAATCAATAGTGAATATGATAAAATGATAATGAAATTAAGTATTACCTGAT 1230
Db 539 TTTTATTAATCAATAGTGAATATGATAAAATGATAATGAAATTAAGTATTACCTGAT 598

QY 1231 TTGGCTTCGATTCGACACCGGCAACAATTCATGACGAAATGTTTATATATAAGAC 1290
Db 599 TTGACCGCGATTTCTAATACCTCTACTAATTTACGTGACGAGCTATATATATAGTAT 658

QY 1291 CACTTATCCGAATGTTGAATAAATTTGATCAATCAATAGTGGGAGAGATAGTCAGAA 1350
Db 659 CATTATCCGAAGTAA-----TTAA 680

QY 1351 ATAAAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1409
Db 681 AAAAAGCAGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 740

QY 1410 TTGCATCATCTTTTGTATATATCTCAAAAGATCTCTTGTTCGCAATAGTCAATCTTC 1469
Db 741 TTGCATCATCTTTTGTATATATCTCAAAAGATCTCTTGTTCGCAATAGTCAATCTTC 800

```

LOCUS A0502594 537 bp DNA linear GSS 29-APR-1999
DEFINITION V5287 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
genomic 5', genomic survey sequence.
ACCESSION A0502594
VERSION A0502594.1 GI:4705140
KEYWORDS GSS.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 537)
AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,
desRoches, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R.,
Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
Gene Disruption
JOURNAL Unpublished (1999)
COMMENT Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mTn-3xHA/lacZ insertion.
Seq primer: GGCCTTCTTCTTTGGAAGTAC
Class: transposon-tagged.
Location/Qualifiers
1..537
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
/lab_host="E. coli"
/clone_lib="mTn-3xHA/lacZ Insertion Library"
/note="vector: PHSS6-Sal; A yeast genomic DNA library
(lacking mitochondrial DNA) was prepared in PHSS6-Sal;
genomic DNA was size-fractionated (DNA of roughly 2-3 kb
in length) prior to cloning. This library was
subsequently mutagenized with a mTn-3xHA/lacZ
minitransposon containing lacZ, URA3, and tet resistance."

misc_feature complement(<649..>939)
/note="similar to Saccharomyces cerevisiae ORF YPL070w [
weak similarity to Vps9p]"
/evidence=not_experimental
ORIGIN
Query Match 29.8%; Score 458.2; DB 9; Length 1021;
Best Local Similarity 76.1%; Pred. No. 7.3e-102;
Matches 649; Conservative 0; Mismatches 158; Indels 46; Gaps 4;
QY 724 CAAGCTTGGATATATCTAGGAGAGCTTTCTGCTGAAATCATATCACTACTCAGAGATG 783
DB 1 CAAGGTTGGATATATCTAGGAGAGATTTTCTACCTGAAATCATATCACTACTCAGAGATG 60
QY 784 TATTTGAATATGTTATGAATAAACAAGCGGCGCTTTTCAGATTAACGGTGGAGCTCATG 843
DB 61 TATTTGAATATGTTATGAATAAACAAGCGGCTGCTTTTCAGATTAACGGTGGAGCTCATG 120
QY 844 GAAGCGCTGTCCTTCTTCCACACGCGGCGATTCGTTGGTTCCTTTTCATAATCTTCG 903
DB 121 GAAACGTTCTCTCTTCTTCCACACGCGGCGATTCGTTGGTTCCTTTTCATAATCTTCG 180
QY 904 GGTATTTATTCAGATTAAGATGATTAATCTTGAATTTGAAAGATTTCCAAATGTCCAGC 963
DB 181 GGTATGATTTATCAGATTAAGATGATTAATCTTGAATTTGAAAGATTTCCAAATGTCCAGC 240
QY 964 GAAAAGGCTTTGCTGAGGACATTAACAGAGGGGAAGTTATCTTTTCCGATCGTCCAGCC 1023
DB 241 GAAAAGGCTTGGCTGAGGACATTAACAGG -AGGAAGCTCTCTTTTCCGATCGTCCAGCC 299
QY 1024 CTTAATCTTCACTAATAAGAGTCTAATCTGACGACACCAATCTTGAATTTCTC 1083
DB 300 CTAAATTTCCACCAAGCGCGAAGCTCAAACTGAGCAACACGATGAATTCATGAATTTCT 359
QY 1084 CTCTTGAGGACAACTGATTAAGATATAAACTAAAGCTGATTCAAATCTGGAATTCGAC 1143
DB 360 TTAATGAGGACTAATGATACAGCTTAATACTGAAATGATTAATCTTGAATTTCCAG 419
QY 1144 ACCAATTTCAATGGCTTACACCAAAATTTTATTAATCAATAGTGAATGATAAATAAT 1203
DB 420 ACCAATTTCCCTGAATATAGCAAAAGATTTTATCAATCAGTTAGTGAACATGATAAATAAT 479
QY 1204 GATAATGAATAATAGTATTTACCTGATTTGCTTCCGATTCGACACCGCCACCAATTTA 1263
DB 480 GATGACAAAATAAGTATTTGCGCGATTTGACCCGCTTCTTAATCTTCACTAATTTA 539
QY 1264 CATGACGAATGTTATATAATAGACCACTTATCCGAATTTGTCAATAAATGATCAAT 1323
DB 540 CGTGACGAGCTATATATATAGTAGATCAATTTATCCGAACTGTAA ----- 584
QY 1324 CAAATTAGTGAGGAGATAGTACGAATAAAGCTTCTCTCTCTCTCTCTCTCTCTCTCTAT 1383
DB 585 -----TTAAAAAAGCAGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAT 621
QY 1384 ACATAC -GATTTCAATATAGTTTTCATCTGATCATCTTTTGTATATATCTCAAAAAGATC 1442
DB 622 ATATACAGTATTTATTTATAGTATATTTGATCATCTTTTGGTATTTTCAATAATC 681
QY 1443 TCTTAGTTTCGAAATAGTCAAAATCTTCAAAATTTATAGCTTTATTTTTCACGATTT 1502
DB 682 TTTTAGTTTCGATATAGTCAAAATCTTCAAAATTTGATTTTTCACGATTT 741
QY 1503 CTGAAATCTCTTTTATCAGACCGTTAATGC -----TAGCGGTTACTGTCAATCGCC 1556
DB 742 TTGGATTTCTCTCTCAATACCGCTAACACTTGCATTTGGCGGCTACTGTCAATGACC 801
QY 1557 GGTAAATTCGGA 1569
DB 802 TGTAAATTCAGA 814
RESULT 4
A0502594/c

```

Db      151 TTTTAAATCGACGATATAGAAGATAATGCTCCCTTGAGAGGGGACAGACCACTTCTCACT 92
QY      575 TAATCTTCGGTGTACCTCCACTATAA 601
        |||||
Db      91 TAATCTTCGGTGTACCTCCACTATAA 65

RESULT 5
CNS06CKP      1115 bp      DNA      linear      GSS 14-JUN-2001
LOCUS      T7 end of clone AR0AA006A06 of library AR0AA from strain CBS 732 of
DEFINITION      Zygosaccharomyces rouxii, genomic survey sequence.
ACCESSION      AL392639
VERSION      AL392639.1 GI:12142195
KEYWORDS      GSS.
SOURCE      Zygosaccharomyces rouxii
ORGANISM      Zygosaccharomyces rouxii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes.
            1 (bases 1 to 1115)
            Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Liorente,B.,
            Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
            Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
            FEMS Lett. 487 (1), 3-12 (2000)
            20584711
            11152876
            2 (bases 1 to 1115)
            de Montigny,J., Straub,M., Potier,S., Tekala,F., Dujon,B.,
            Wincker,P., Artiguenave,F. and Souciet,J.
            Genomic exploration of the hemiascomycetous yeasts: 8.
            Zygosaccharomyces rouxii
            FEMS Lett. 487 (1), 52-55 (2000)
            20584718
            11152883
            3 (bases 1 to 1115)
            Genoscope.
            Direct Submission
            Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            sege@genoscope.cns.fr - Web : www.genoscope.cns.fr)
            This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
            Location/Qualifiers
            1..1115
            /organism="Zygosaccharomyces rouxii"
            /mol_type="genomic DNA"
            /strain="CBS 732"
            /db_xref="taxon:4956"
            /clone="AR0AA006A06"
            /clone_lib="AR0AA"
            /note="end : T7"
            <2..>778
            /note="similar to Saccharomyces cerevisiae ORF YPL069c [
            BT51 ; geranylgeranyl diphosphate synthase ]"
            /evidence=not_experimental
            complement(<943..>1020)
            /note="similar to Saccharomyces cerevisiae ORF YPL070w [
            weak similarity to Vps9p ]"
            /evidence=not_experimental

FEATURES
            source
            misc_feature
            misc_feature

ORIGIN
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Query Match      12.4%; Score 195; DB 9; Length 1115;
Best Local Similarity 55.7%; Pred. No. 5.2e-36;
Matches 444; Conservative 0; Mismatches 335; Indels 18; Gaps 3;

QY      420 AAATTTAATAGTTCAAAATTAACAGAGTTATGAAATTTGCCCAAGACCAGCTGGCCATAGT 479
        |||||
Db      1 AAATTTAATAGTTGTTTAAAGAAATATATCAAAATCAAAATTTGAGAAGACCAATCAAAAT 60

QY      480 TTGCGAAATTTGTAGAGCTCTTTCATATAATTCAGACCTTTTAAATCGACGATATAGAAGATAA 539
        |||||
Db      61 CTCACAGTTTAGTGGCCACTTACACAAATGCAAGCTTATTAATAGATGACATTTGAAGATAG 120

QY      540 TGCTCCCTTGAGAGGGGACAGACCACTTCTCATTAAATCTTCGGTGTACCTCCACTAT 599
        |||||
Db      121 CTCGCTACAAGAGAGGCATACCGACTTCATACGTTGTTTGGAGTTCCGCTAACGAT 180

QY      600 AAACACCCCAATATATATGTTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAA 659
        |||||
Db      181 AAATGCTCAAAATTCATAGTATTTTAAAGCTATGGAATTTCTGCAACGATTTCTGGTGA 240

QY      660 AG---AGCCTTTGTATCAATAATTTGATTACGATTTTCAACGAAAGAAATTCATCAATCTACA 716
        |||||
Db      241 TGACAAAATTTCTGTACTCGAATTAATGTTTATCTTCAATCAGGAAATGATTAATTTGCA 300

QY      717 TAGGGACAAGCGTTGGATATATATCTGGAGAGACTTTCTGCTGTAATCATACCTACTCA 776
        |||||
Db      301 TAGAGGCCAAGTTTAGATATTTACTGCGGGATAAATTTATTGCGATTTATACCAAGATGA 360

QY      777 GGAGATGTTATTGAATATGTTTATGAATAAACAGGGCGCCTTTTCAGATTAACGTTGAG 836
        |||||
Db      361 AGCAGCTACTATATATATGTTGATGAACAGACTGGTGGCTTATTTAGATTAAACGGTTAG 420

QY      837 ACTCATGGAAGCGCTGTCTCCTCTCACACCCAGCGCATTCGTTGGTTCCTTTCAVAAA 896
        |||||
Db      421 AATCATGGAATTTGCTGACCGAT---GTAGATTATCCGCAATTCATTAGTACCGCTGAGCA 477

QY      897 TCTTCTGGGTATTTATTCAGATTAGAGATGATTCTTGAATTTGAAAGATTTCCAAAT 956
        |||||
Db      478 CTTTTTGGGAATAATTTATCAAAATTCGCGATGATTATCAAAATCTTTTCAATGAACAAAT 537

QY      957 GTCCAGCGAAAAAGCGTTTGTCTGAGGACATTACAGAGGGGAGTATCTTTTCCATCGT 1016
        |||||
Db      538 GATAGCTAATTAAGGTTTAGCAGAAGATATTTCTGAAGGTAAACTATCTTTCCCGATTAT 597

QY      1017 CCACGCCCTTAACTTCACTTAAACGAAAGGTCAAACTGAGCAACACAAATGAAATTTCTAAG 1076
        |||||
Db      598 CCACGGCTTGAAT-----ATGCTCAATCTCAAAACGAAACTATTTTGCAGA 645

QY      1077 AATTCTCTGTGAGGACAAGTATTAAGATATAAACTAAAGCTGATTCAAATACTGGA 1136
        |||||
Db      646 CGTTCTACGGTTAGCAACAGATGATGTTGAATGAAATGAAAAAAGGGCATAGACTACCTAA 705

QY      1137 ATTGCAACCAATTCATTTGCCCTACACCAAAATTTTATTAATCAATAGTATATAT 1196
        |||||
Db      706 TAACGAAGTCACTCCCTAGATGACACTTGGGAAAAGCTCGAGAACTCTCAAAATTTGCG 765

QY      1197 AAAAAATGATAATGAAA 1213
        |||||
Db      766 TAAATCAGAGAATTATA 782

RESULT 6
AY407652      903 bp      DNA      linear      GSS 15-DEC-2003
LOCUS      Homo sapiens GGPS1 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION      genomic survey sequence.
ACCESSION      AY407652
VERSION      AY407652.1 GI:39763623
KEYWORDS      GSS.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 903)
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferrera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
Infering nomenclature evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
14671302
2 (bases 1 to 903)
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferrera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
Location/Qualifiers
1..903
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>903
/gene="GGPS1"
/locus_tag="HCM2962"
ORIGIN
Query Match 7.3%; Score 114; DB 9; Length 903;
Best Local Similarity 52.1%; Pred. No. 1.7e-16;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;
QY 359 AAGCTTGATTTCACAACTTATAATCACATCCTTTTGAACCTGCGCAAGACTTTAGAC 418
DB 26 AAGAATTCTTCAGAACCTTATAATCTTCTCAGTTACCGAGTAAACAGTACAGAA 85
QY 419 TAAATTAATAGTTCAAAATTAACAGAGTATGATTTTGCCAAAGCAGCTGGCCATAG 478
DB 86 CCAAACTTTCACAGGCAATTAATCAATTTGGCTGAAAGTTCCAGAGGCAAGCTACAGATTA 145
QY 479 TTTCGCAATTTGTTAGCTCTCCATAATTCACAGCTTTTAACTCGACGATATAGAAGATA 538
DB 146 TTAATGAAGTGACAGAAAGTTGCGAATGCCAGTTTACTCATCGATGATTTGAAGACA 205
QY 539 ATGCTCCCTTGAGAAAGGGGACAGACCACCTTCTCACTTAATTCCTCGGTGTACCTCCACTA 598
DB 206 ACTCAAACTCCGACGTTGGCTTTCCAGTGGCCACAGCATCTATGGAATCCCATCTGTCA 265
QY 599 TAAACACCGCAATATATGTTTTCAGAGCCATGCACTGTGTATCGCAGCTAACACAA 658
DB 266 TCAATTCGCAATTAAGTGTATTTCTTCTGGCTGGAGAAAGTCT----- 310
QY 659 AAGAGCCTTTGTATCAATAATTTGATTTACGATTTTCAACGAAGATTTGATCAATCTACATA 718
DB 311 TAACCTTGATACCCAGATGCGAGTATTTTACCCGCGAGCTTTTGGAACTCCATC 370
QY 719 GGGGACAAGGCTTGGATATATCTACTGAGAGACTTTCTGCTCGAATCATACCTACTCAGG 778
DB 371 AGGACAAGGCTTAGATATTTACTGGAGGA-----TAAATACATCTTGCTCCCACTGAAG 424
QY 779 AGATGATTTGAATATGTTATGATTAATAAACAAGCGGCGCTTTTCAGATTAACGTTGAGAC 838
DB 425 AAGAATATAAAGCTATGTTGTCGAGAAAACAGGTGGAGCTGTTTGGATTTAGAGTAGTC 484
QY 839 TCATGGAAGCGCTGCTCTCTCTCCCTCACACCAACGCGCAATTCGTTGGTTTCTTTTCAATAATC 898
DB 485 TCATGCAAGTGTCTCTGATTTACAAAGAGATTTAAACCGCTACTT-----AATA 535
QY 899 TTCTGGGTATTTATATACAGATTAAGATGATTTCTGAAATTTGAAAGATTTCCAAATGT 958
DB 536 CACTTGGGCTCTTTTCCAAATTAGGATGATTTATGCTAATCTACACTCCAAAGATATA 595

QY 959 CCAGCGAAAAAGGCTTTCTGAGGACATTCAGAGGGGAGGAGTATCTTTTCCATCGTCC 1018
DB 596 GTGAAAAACAAGTTTGTGAGAGATCTGACAGAGGGAAGTCTCTCAATTCCTACTATTC 655
QY 1019 AGCGCCCTT 1026
DB 656 ATGCTATT 663
RESULT 7
LOCUS BX460174 1076 bp mRNA linear EST 06-MAY-2004
DEFINITION BX460174 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CS0DF011YP23 5-PRIME, mRNA sequence.
ACCESSION BX460174 GI:47072855
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1076)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 22, 2003 this sequence version replaced gi:31037026.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 868.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DF011CH12QF1&c=868.f.
FEATURES
Location/Qualifiers
1..1076
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF011YP23"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/notes="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
ORIGIN
Query Match 7.3%; Score 114; DB 5; Length 1076;
Best Local Similarity 52.1%; Pred. No. 1.7e-16;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;
QY 359 AAGCTTGATTTCAAACTTATAATCACATCCTTTTGAACCTGCGCAAGACTTTAGAC 418
DB 251 AAGAATTTCTTAGAACCTTATAATCAATTTGGCTGAAAGTTCCAGAGGCAAGCTACAGATTA 310
QY 419 TAAATTTAATAGTTCAAAATTAACAGAGTATGATTTTGCCCAAGACCACTGGCCATAG 478
DB 311 CCAAACTTTCACAGGCAATTAATCAATTTGGCTGAAAGTTCCAGAGGCAAGCTACAGATTA 370
QY 479 TTTCGCAATTTGTTAGCTCTTGCATTAATTCAGCCCTTTTAAATCGAGATATAGAATA 538
DB 371 TTATTTGAAGTCACAGAAATGTTGCATAATGCCAGTTTACTCATCGATGATTTGAAGACA 430
QY 539 ATGCTCCCTTGAGAGGGGACAGACCACCTTCTCACTTAATCTTCGCTGTACCTCCACTA 598

```

Db      431 ACTCAAACTCCGACGTGGCTTCCAGTGGCCGACACAGCATCTATGGAATCCCATCTGTCA 490
QY      599 TAAACACCGCAATATATATGTTTTCAGAGCCATGCAACTTGTATCGCAGCTAACACAA 658
Db      491 TCAATTTCTGCCAATACGTGTATTTCTTGGCTTGGAGAAAGTCT-----535
QY      659 AAGAGCTTTGATCATAAATTTGATACGATTTTCAACGAAGAATTTGATCAATCTACATA 718
Db      536 TAAACCTTTGATCACCCAGATGCAAGTCTTTTACCCGCCAGCTTTTGGAACTCCATC 595
QY      719 GGGGCAAGGCTTGGATATATATGTTTTCAGAGCCATGCTTCTGCTGAAATCATCTACTCAG 778
Db      596 AGGCAAGAGCCCTAGATATTTTACTGGAGGA-----TAAATTAACATTTGTCCCACTGAAG 649
QY      779 AGATGATTTTGAATATGTTTATGAATAAACAAGCGGCTTTTCAGATTAAGCTTGGAGAC 838
Db      650 AAGAATAATAAGCTATGCTGCTGCAGAAAACAGGTGGTGTGATTTAGCAGTAGGTC 709
QY      839 TCATGGAAGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 898
Db      710 TCATGCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 760
QY      899 TTCTGGGTATTTATTTATCAGATTTAGAGATTTACTTGAATTTGAAAGATTTCCAAATGT 958
Db      761 CACTTGGGCTCTTTTCCAAATTTAGGATGATTTATGCTTAATCTACACTCCAAAGATATA 820
QY      959 CCAGCGAAAAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1018
Db      821 GTGAAAACAAAGTTTTTGTGAGATCTGACAGAGGGAAGTTCTCAATTTCTACTATTC 880
QY      1019 ACGCCCTT 1026
Db      881 ATGCTATT 888

RESULT 8
LOCUS   CR612465
DEFINITION full-length cDNA clone CS0D1051YM07 of Placenta Cot 25-normalized
of Homo sapiens (human).
ACCESSION CR612465
VERSION   1 GI:50493272
KEYWORDS  HTC; CNSLT_CDNA.
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
JOURNAL   Full-length cDNA libraries and normalization
REMARK    Unpublished
Contact : Feng Liang Email : fliang@lifetech.com URL :
Paraday Avenue
Genoscope.
2 (bases 1 to 1417)
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT  1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source      Location/Qualifiers
1..1417
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1051YM07"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

```

ORIGIN

```

Query Match      7.3%; Score 114; DB 3; Length 1417;
Best Local Similarity 52.1%; Pred. No. 1.8e-16;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;

QY      359 AAAGCTTGTATTTCAAAACCTTATATATCAATCTTTTGAACCTGGCAAGAACTTTAGAC 418
Db      249 AAAGAATTTCTTGAAGACCTTATAAATATCTTACTTTCAGTTACCAAGTAAACAGTGAGAA 308
QY      419 TAAATTTAATAGTTTCAAAATTAACAGAGTTATGAAATTTGCCAAAGACAGCTGGCCATAG 478
Db      309 CCATATTTTCAAGCAATTTAATCATTTGGCTGAAAGTTCCAGAGCAAGCTACAGATTA 368
QY      479 TTTCGCAATTTGTGAGCTCTTGCATAATTTCCAGCTTTTATATCGACGATATAGAGATA 538
Db      369 TTATTGAAGTGCAGAAATTTGCATAATGCGAGTTTACTCATCGATGATATGAGACA 428
QY      539 ATGCTCCCTTGAAGAGGGGACAGACCACTTCTCACTTAATCTTCTGTTGTTACCTCCACTA 598
Db      429 ACTCAAACTCCGACGTGGCTTTCCAGTGGCCACAGCATCTATGGAATCCCATCTGTCA 488
QY      599 TAAACACCGCAAAATATATGTTTTCAGAGCCATGCAACTTGTATCGCAGCTTAACCAAA 658
Db      489 TCAATTTCTGCCAATTAAGTGTATTTCTTGGCTTGGAGAAAGTCT-----533
QY      659 AAGAGCTTTGTATCATAAATTTGATTTACGATTTTCAACGAAGAATTTGATCAATCTACATA 718
Db      534 TAAACCTTTGATCACCCAGATGCAAGTCTTTTACCCGCCAGCTTTTGGAACTCCATC 593
QY      719 GGGGCAAGGCTTGGATATATATGTTTTCAGAGCCATGCTTCTGCTGAAATCATCTACTCAG 778
Db      594 AGGCAAGAGCCCTAGATTTTACTGGAGGA-----TAAATTAACATTTGTCCCACTGAAG 647
QY      779 AGATGATTTTGAATATGTTTATGAATAAACAAGCGGCTTTTCAGATTAAGCTTGGAGAC 838
Db      648 AAGAATAATAAGCTATGTTGCTGCAGAAAACAGGTGGTGTGATTTAGCAGTAGGTC 707
QY      839 TCATGGAAGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 898
Db      708 TCATGCACTGTTCTCTCTGATTACAAAGAGATTTAAACCCGCTACTT-----ATA 758
QY      899 TTCTGGGTATTTATTTATCAGATTTAGAGATTTACTTGAATTTGAAAGATTTCCAAATGT 958
Db      759 CACTTGGGCTCTTTTCCAAATTTAGGATGATTTATGCTTAATCTACACTCCAAAGATATA 818
QY      959 CCAGCGAAAAGGCTTGTGCTGAGACATTACAGAGGGGAGTATCTTTTCCCATCTCTCC 1018
Db      819 GTGAAAACAAAGTTTTTGTGAGATCTGACAGAGGGAAGTTCTCAATTTCTACTATTC 878
QY      1019 ACGCCCTT 1026
Db      879 ATGCTATT 886

```

RESULT 9

```

LOCUS   CR603114
DEFINITION full-length cDNA clone CS0D001VK22 of T cells (Jurkat cell line)
of Homo sapiens (human).
ACCESSION CR603114
VERSION   1 GI:50483921
KEYWORDS  HTC; CNSLT_CDNA.
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
JOURNAL   Full-length cDNA libraries and normalization
REMARK    Unpublished
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

```

Faraday Avenue
2 (bases 1 to 2703)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segrefgenoscope.cns.fr
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
source
1..2703
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ001YK22"
/tissue_type="T cells (Jurkat cell line) Cot
10-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 7.3%; Score 114; DB 3; Length 2703;
Best Local Similarity 52.1%; Pred. No. 2e-16;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;
QY 359 AAGCTTGATTTCAAAACCTTATATACATCCCTTTTGAACCTGCGAAGACTTTAGAC 418
Db |||||
1533 AAGAAGTCTCTAGAACCTTATAATCTTACTTACGTTACCGTTAAACAGTGAGAA 1592
QY 419 TAAATTTAATAGTTCAATTTACAGAGTTATGAATTTGCCAAAGACAGCTGGCCATAG 478
Db |||||
1593 CCAAACTTTCACAGGCAATTAATCATTTGGCTGAAGTTCCAGAGACAGCTACAGATTA 1652
QY 479 TTTGCGAAATTTGTGAGCTCTTGCATPAATTTCCAGCCTTTTAAATCGACATNTAGAGATA 538
Db |||||
1653 TTATTCAGTGACAGAAATGTTGCAATATGCCAGTTTACTCATCGATGATATTGAAGACA 1712
QY 539 ATGCTCCCTTGAAGAGGGACAGACCACTTCTCACTTATCTTCGTTGACCTCCACTA 598
Db |||||
1713 ACTCAAACTCCGACGTGGCTTTCCAGTGGCCACACAGCATCTATGGAATCCCATCTGTCA 1772
QY 599 TAAACACCCCAATATATATGTTATTCAGAGCCATGCACTTGTATCGCAGTAAACACAA 658
Db |||||
1773 TCAATTTCTGCAATTAACGTGATTTCTTGGCTTGGAGAAAGTCT----- 1817
QY 659 AAGAGCCTTTGTATCATATTTGATTTACGATTTTCAACGAGAAATGATCAATCTACATA 718
Db |||||
1818 TAACCTTGTATCACCAGATGCAATGAGCTTTTACCCGCCAGCTTTTGGAACTCCATC 1877
QY 719 GGGGCAAGGCTTGGATATATCTGAGAGACTTTCTGCTGTAATCATACCTACTCAGG 778
Db |||||
1878 AGGGAACAAGCCCTAGATATTTACTTGGAGGA-----TAATACATGTGCCACTGAG 1931
QY 779 AGATGTATTTGAATATGTTTATCAATAAACAAGCGCGCCCTTTTCAGATTAACCTTCAGAC 838
Db |||||
1932 AGAATATAAGTAATGTTGCTGCAAGAAACAGGTGGACTGTTTGGATAGCAGTAGGTC 1991
QY 839 TCATGAAGCGCTGTCTCTCTCTCTCTCAACACCGGCCATTCGTTGGTTTCTTTTCAATAATC 898
Db |||||
1992 TCATGAGTGTGTTCTCTGATTACAAGAGAGATTTAAACCGGTACTT-----AATA 2042
QY 899 TTTGCGGTATTTATTCAGATTTAGATGATTTACTTGAATTTGAAGATTTTCAATATGT 958
Db |||||
2043 CACTTGGCTCTTTTTCAAATTTAGGAGATTTATGCTAATCTACATCTCAAGAAATATA 2102
QY 959 CCAGCGAAAGGCTTTGCTGAGCAATTACAGAGGGGAAGTTATCTTTTCCCATCTGCC 1018
Db |||||
2103 GTGAACAACAAGTTTTTGTGAGATCTGACAGAGGGAAGTTTCTCATTTCTTACTATT 2162
QY 1019 ACGCCCTT 1026
|||

Db 2163 ATGCTATT 2170
RESULT 10
AY407654 903 bp DNA linear GSS 15-DEC-2003
LOCUS Mus musculus GBS1 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION Genomic survey sequence.
ACCESSION AY407654
VERSION AY407654.1 GI:39763625
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 903)
REFERENCE Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
AUTHORS Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 903)
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
source
1..903
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>903
/gene="GSP1"
/locus_tag="HCM2962"
ORIGIN
Query Match 7.2%; Score 113.6; DB 9; Length 903;
Best Local Similarity 51.9%; Pred. No. 2.1e-16;
Matches 350; Conservative 0; Mismatches 294; Indels 30; Gaps 3;
QY 353 AAAATGAAGCTTGATTTCAAAACCTTATAATCACATCCTTTTGAACCTGCGCAAGACT 412
Db 20 AAGCTGAGAGGATCTCTTAGAGCCCTATAGGTACTTACTTACGTTACAGGTTAAACAGG 79
QY 413 TTAGACTAATTTAATAGTTCAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGG 472
Db 80 TGGAAGCAACTTTTCACAGGCATTTAATCACTGGCTGAAAGTTCCAGAGACAAGCTAC 139
QY 473 CCATAGTTTTCGAAATTTGTGAGCTCTTGCAATTTCCAGCCTTTTAAATCGACATATAG 532
Db 140 AGATTATCATTTGAAGTGACTGAAATGTTGCATAATGCCAGTTTACTCATTTGATGATTTG 199
QY 533 AGATTAATGCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGTTGTTACCT 592
Db 200 AAGACAGTTTCAAAGCTCCGACGTTTCCAGTGGCTACAGCATCTATGTTGTTCCCAT 259
QY 593 CCACATAACACCGCAATTTATATGTTTACAGAGCCATGCAACTTTGTTATCGACGTAA 652
Db 260 CTGTCTAATTTCTGCCAATTACGTCTACTTCTTGGACTGGAAAAAGTCT----- 310
QY 653 CCACAAAGAGCCTTTGTTATCATTAATTTGATTTAGGATTTTCAACGAGAGATTTGATCATC 712
Db 311 -----TAACCTTTGATCACCCGGATGCGGTGAAGCTTTTTCACACGCGACGCTTCTGGAAC 364
QY 713 TACATAGGGGACAGGCTTGGATATATCTGAGAGACTTTTCTGCTGAAATCATACCTA 772

444 TGAGAAGCAAACTTTCACAGSCATTTATCACTGGCTGAAAGTTCCAGAGACAAAGCTAC 503
473 CCATAGTTTCGCAAAATGTTTTCAGCTCTTTCGCTAATATTCAGCCTTTTATTCACAGATATAG 532
504 AGATTATCATTTGAAGTGAAGTGAATGTTTGCATAATGCGCAGTTTACTCATTTGATGATATTTG 563
533 AAGTAATATGCTCCCTTTCAGAGGGGACAGACCACTTCTCACTTATCTTCGGTGTGACCT 592
564 AAGACAGTTCAAAGCTCCGAGTGGTTCCTCCAGTGGCTTCACAGCACTATGTTGGTGTCCCAT 623
593 CCACATATAACACCCCAATATATATGATTTTTCAGAGCCATGCAACTTGTATGCGAGTAA 652
624 CTGTCATTTAATCTGCCAATTAAGTCTACTTCTTGGAGTGAAGAGTCT----- 674
653 CCACAAAAGAGCCTTTGTATCAATTAATTTGATTAACGATTTTTCACGAGAAATGATCAATC 712
675 -----TAACCTTTCATACCCGGATGCGGTGAAGCTTTTTCACGCGCAGCTTCTGGAAC 728
713 TACATAGGGGACAAAGCTTGGATATATCTGAGAGACTTCTGCTGAAATCATACCTTA 772
729 TTTCACTAGGAGCAAGGCTTCGATATTTTCTGAGGAGC-----ACCTACACTTGTCCAA 782
773 CTCAGAGATGATTTGAATATGTTTATGAATAAAGCAAGCGGCGCTTTTCAGATTAACGT 832
783 CTGAAGAAGAAATATAAGCCATGTTTTCAGAGACAGTGGTGTGTTGGATAGCAG 842
833 TGAGACTGAGGAAGCGCTGCTCTCTCTCTCTCAACAGCGGCATTCGTTGGTTCCTTTCA 892
843 TAGGTCTTATGACAGCTGTTCTCTGATTTACAAAGAGATCTAAAGCCACCTGCTTG----- 896
893 TAAATCTTCTGGGTATTTATCATAGATGAGATGATTTGATTTGAAAGATTTTCC 952
897 ---ACACACTTGGGCTTTTTCAGATGAGATGATTTGCAATCTACACTCCAAAG 953
953 AAATGTCCAGCAAAAAGCTTTGCTGAGGACATTAACAGAGGAGATTTATCTTTTCCCA 1012
954 AATACAGTGAACAAGATTTCTGTAAGACTTGACAGAGGAGATTTCTCATTTCCCA 1013
1013 TCGTCCAGCGCTT 1026
1014 CTATCCATGCCATT 1027

RESULT 12
BI520491
LOCUS
DEFINITION
603071690F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5163599 5',
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 815)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L14M11406 Row: c Column: 24
High quality sequence stop: 813.
Location/Qualifiers
1..815
/organism="Homo sapiens"

FEATURES
source

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5163599"
/tissue_type="medulla"
/lab_host="DH10B"
/clone_lib="NIH_MGC_119"
/note="Organ: brain; Vector: pCMV-SPORT6; Site: 1; NotI;
Site 2: EcoRV (destroyed); RNA source normal medulla from
anonymous male age 27. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.3 kb, insert size range
0.9-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 013. Note:
this is a NIH_MGC Library."

ORIGIN

Query Match 6.6%; Score 103; DB 4; Length 815;
Best Local Similarity 52.0%; Pred. No. 7.5e-14; Indels 31; Gaps 4;
Matches 348; Conservative 0; Mismatches 290;
359 AAAGCTTGATTTCAAAACCTTATAATCACATCTTTTGAACCTGCGCAAGAACTTTAGAC 418
113 AAAGAATTTCTTACAGCCCTATATAATCTTACTTACGTTACCGGTAAACAAGTCAGAA 172
419 TAAATTTAATAGTTCAAATTAACAGAGTTATGAAATTTGCCAAAGACCGAGCTGGCCATAG 478
173 CCAAACTTTACAGGCAATTTAATCATTTGGCTGAAAGTTCCAGAGGACCAAGCTACAGATTA 232
479 TTTCCGAAATTTTTCAGCTCTTTCAGATTAATCCAGCTTTTAAATCCAGCATATAGAAGATA 538
233 TTTAAGAGTGACAGAAATGTTTCATATGCGCAGTTTACTCATCATGATATTTGAAGACA 292
539 ATGCTCCCTTGAAGAGGGGACAGACCACTTCTCACTTATCTTCGGTGTACCTCCACATTA 598
293 ACTCAAACTCCGACGTCGCTTTCCAGTGGCCACAGCATCTATGAAATCCCATCTGTCA 352
599 TAAACAACCCAAATTTATATATTTTCAGAGCCATGCAACTTGTATCGAGCTAACACAA 658
353 TCAATTTCTGCAATTAACGTGTATTTTCTTGGCTTGGAGAAAGTCT----- 397
659 AAGAGCTTTGTATCATATAATTTGATTAACGATTTTCAACGAAGAAATTTGATCAATCTACATA 718
398 TAACTCTTGAATCAACCGATGAGTGAAGCTTTTACCGCCAGCTTTTGGAACTCCATC 457
719 GGGGACAAAGCTTGGATATATATCTGAGAGACTTTCTGCTGAAATCATACCTACCTCAGG 778
458 AGGACAAAGGCTTAGATATTTACTGAGGGA-----TAATTACACTTGTCCCACTGAAG 511
779 AGATGATTTGAATATGTTTATGAATAAACAAGCGGCGCTTTTTCAGATTAACCTTGAGAC 838
512 AAGATATAAAGCTATGTTGCTGCTGAGAAACAGGTGGAGCTGTTTGGATTTAGCAGTAGGTC 571
839 TCATGGAAGCGCTGCTCTCTCTCTCCACACCGCCATTCGTTGGTTCCTTTTCATTAATC 898
572 TCATGAGTGTGTTCTCTGATTACAAGAAGATTTAAACCGCTACTT-----AATA 622
899 TTTCTGGTATTTATTTATTCAGATTTAGATGATTTGAAATTTCAAGATTTCCAAATGT 958
623 CACTTGGGCTCTTTTCCAAATTTAGGATGATTTATGCTAATCTACACTCCAAAGAAATATA 682
959 CCAGCGAAAGAGCTTTGCTGAGGACATTTACAGAGGGG- GAAGTTATCTTTTCCCATCGTC 1017
683 GTGAAAAACAAAAGTTTGTGAAGATCTGACAGAGGGCAAGTTCTCATTTCCTACTATT 742
1018 CAGCCCTT 1026
743 CATGCTATT 751

RESULT 13
BX439729
LOCUS

BX439729 905 bp mRNA linear EST 04-MAY-2004

DEFINITION BX439729 Homo sapiens PLACENTA Homo sapiens cDNA clone CSODE014Y115
 5-PRIME, mRNA sequence.
 ACCESSION BX439729
 VERSION BX439729.2 GI:47012615
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 905)
 AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT On May 15, 2003 this sequence version replaced gi:30785818.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
 was not normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.
 This sequence belongs to sequence cluster 868.f
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?s=CSODE014AE08QP1&c=868.f.
 http://www.genoscope.cns.fr/cdna?s=CSODE014AE08QP1&c=868.f.

FEATURES

source
 1. .905
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CSODE014Y115"
 /tissue type="PLACENTA"
 /clone_lib="Homo sapiens PLACENTA"
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

ORIGIN

Query Match 6.3%; Score 99; DB 5; Length 905;
 Best Local Similarity 51.8%; Pred. No. 7e-13;
 Matches 338; Conservative 2; Mismatches 282; Indels 31; Gaps 4;

QY 359 AAAGCTTGATTCAAAACCTTATATATACATCCCTTTTGAACCTGGCAAGACTTTAGAC 418
 DB |||||
 QY 209 AAAGAATCTCTAGAACCTTATAAATCTTACTTCAGTACCAGGTAAACAAAGTGAGAA 268
 DB |||||
 QY 419 TAAATTTAATAGTTCAATTAACAGAGTTATGAAATTTGCCCAAGACGAGCTGGCCATAG 478
 DB |||||
 QY 269 CCAAACTTTCACAGGCATTTAATCATTTGGCTGAAAGTTCAGAGGACAAGCTACAGATTA 328
 DB |||||
 QY 479 TTTTCGCAATTTGTGAGCTCTTGATTAATTCAGCCCTTTTAAATCGAGATATAGAAGATA 538
 DB |||||
 QY 329 TTAATTGAAGTGACAGAAATGTGCATATATGCGAGTTTACTCATCGATGATATGAAGACA 388
 DB |||||
 QY 539 ATGCTCCCTTGAAGAAGGGGACAGACCACTTCTCAATTAATCTTCGGTGTCACCTCCACTA 598
 DB |||||
 QY 389 ACTCAAACTCCAGCGTGGCTTCCAGTGGCCACAGCATCTATGGAATCCCATCTGTCTCA 448
 DB |||||
 QY 599 TAAACACCGCAATTAATATGATTTTACAGAGCCATGCAACTTGATTCGAGCTAACCCACAA 658
 DB |||||
 QY 449 TCAATTCGCCAATTAAGTGTATTTCTTGGCTTGGGAAGTCT----- 493
 DB |||||
 QY 659 AAGAGCCTTTGTATCATATTTGATACGATTTTCAACGAAGAATTGATCAATCTACATA 718
 DB |||||
 QY 494 TAAACCTTGATCATCCCGAGTGCAGTGAAGCTTTTACCGCCAGCTTTTGGAACTCCATC 553
 DB |||||
 QY 719 GCGGACAGGCTTGGATATATCTGAGGAGACTTTCTGCGCTGAAATCATACCTACTCAGG 778
 DB |||||
 QY 554 AGGACAGAGCCCTAGATATTTACTGGAGGA-----TAATTACATTTGTCCCACTGAAG 607
 DB |||||

QY 779 AGATGTAATTTGAATGTTATGTAATAAAACAGGGCGCCCTTTTCAGATTAAACGTTGAGAC 838
 DB |||||
 QY 608 AAGAATATAAAGCTATATGTTGCTGCAGAAAACAGGTGGACTGTTTGGAATTAGCAGTAGGTC 667
 DB |||||
 QY 839 TCATGGAAGCGCTGCTCTCTCTCCACACCCAGCGCCATTCGTTGGTTCCTTTTCATAAATC 898
 DB |||||
 QY 668 TCAUCAGTGTGTTCTCTGATTACAAAGAGATTAAAACCGTACTT-----AATA 718
 DB |||||
 QY 899 TTCTGGGTATTTATTCAGATTAGAGATGATTACTTGAATTTGAAAGATTTCCTCAATGT 958
 DB |||||
 QY 719 CACTTGGGCTCTTTTCCAAATTTAGGATGATTATGCTAACT-ACATCCCAAGAAATAT 777
 DB |||||
 QY 959 CCAGCGAAAAGGCTTCTCTGAGGACATACAGAGGGAAGTATCTTTTCC 1011
 DB |||||
 QY 778 AGTGAACAAAAGTTTTTGTGAAGATCTGACAGAGGAAGTCTCATTTTCYWC 830
 DB |||||

RESULT 14

BI560099

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LHAM1748 row: m column: 21

High quality sequence stop: 781.

Location/Qualifiers

1. .871

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5295932"

/lab host="DH10B"

/clone_lib="NIH MGC 97"

/notes="Organ: testis; Vector: pBluescriptR (modified

pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI

(gtcgg); Oligo-dT primed using primer

5'-TTTTTTTTTTTTTNN-3', size-selected for average

insert size 2.2 kb and normalized to ROT 5. This is a

primary library enriched for full-length clones and

constructed using the Cap-trapper method (Carninci, in

preparation). Library constructed by M. Brownstein

(NIH/NHGRI, National Institutes of Health). Note: this is

a NIH_MGC Library."

ORIGIN

Query Match 6.2%; Score 97.6; DB 4; Length 871;

Best Local Similarity 53.0%; Pred. No. 1.5e-12;

Matches 269; Conservative 0; Mismatches 219; Indels 20; Gaps 2;

QY 359 AAAGCTTGATTCAAAACCTTATATACATCCCTTTTGAACCTGGCAAGACTTTAGAC 418

DB |||||

QY 203 AAAGAATCTCTTAGAACCTTATAATATTACTTACCAGGTAAACAGTAGAGAA 262

DB |||||

```
QY 419 TAAATTTAATAGTTTCARATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATAG 478
Db 263 CCAAACTTTCACAGGCATTTAATCATNTGGTGAAAGTTCCAGAGCAAGCTACAGATTA 322
QY 479 TTTCGCAATTTGTTGAGCTCTTTGCATAATTTCCAGCTTTTAAATCGACGATATAGAAGATA 538
Db 323 TTATTGAAGTACAGAAATGTTGCATATGCCAGTTTACTCATCGATGATATGAAGACA 382
QY 539 ATGCTCCCTTGAGAAAGGACAGACACTTCTCACTTAATCTTCGGTGTACCTCCACTA 598
Db 383 ACTCAAACTCCGAGCTGGCTTTCCAGTGGCCCAAGCATCTATGGAATCCCATCTGTCA 442
QY 599 TAAACACGCAAAATATATGTTTTCAGAGCCATCAACTTCTGCGAGCTTAACACAA 658
Db 443 TCAATCTGCCAATACGTGTATTTCTTGGCTTGGAGAGCTCT----- 487
QY 659 AAGAGCTTTTGAUCAATATTTGATTAGTATTTCAACGAGAAATTTGATCAATCTACATA 718
Db 488 TAAACCTTGTATCACCAGATGAGTGAAGCTTTTACCGCCAGCTTTTGGAACTCCATC 547
QY 719 GGGGCAAGGCTTGGATATATCTACTGGAGAGACTTTCTGCTGAAATCATACCTACTCAG 778
Db 548 AGGGCAAGGCTTAGATATTTACTGGAGGGA-----TAATTACACTTGTCCCACTGAAG 602
QY 779 AGATCTATTTGAATGTTATGATGAATAAAACAGGGGCGCTTTTCAGATTAAAGTTTGAGAC 838
Db 603 AAGATATAAAGCTATGTTGCTGCAGAAACAGTGGAGCTGTTTGGATTAGCAGTAGTTC 662
QY 839 TCATGGAAGCGCTGTCTCTCTCTCTCACA 866
Db 663 TCATGAGTTGTTCTCTGATTACCAAGA 690
```

```
RESULT 15
BG675075 974 bp mRNA linear EST 01-MAY-2001
LOCUS 602621320F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4746676 5',
DEFINITION mRNA sequence.
ACCESSION BG675075
VERSION BG675075.1 GI:13906471
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 974)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL0595 row: h column: 05
High quality sequence stop: 849.
FEATURES
Location/Qualifiers
1..974
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4746676"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Skn3"
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5Kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
```

```
Query Match 6.1%; Score 96.2; DB 4; Length 974;
Best Local Similarity 51.4%; Pred. No. 3.3e-12;
Matches 320; Conservative 0; Mismatches 273; Indels 30; Gaps 3;
QY 404 GCAAGAACTTTAGACTAAATTTAATAGTTTCAAAATTAACAGAGTTATGAATTTGCCAAG 463
Db 45 GTAAACAAGTGAAGAACCAAACTTTTCACAGGCATTTAATCATTTGGCTGAAGTTCCAGAGG 104
QY 464 ACCAGCTGCCCATAGTTTTCGCAAAATTTGTTGAGCTCTTCGATAATTCAGAGCCCTTTTAAATCG 523
Db 105 ACAAGCTACAGATTTAATTTGAAGTGCAGAAATGTTGATATATGCCAGTTTACTCATCG 164
QY 524 ACGATATAGAAGATAATGCTCCCTTTGAGAAGGGGACAGACCACTTCTCACTTAAATCTTCG 583
Db 165 ATGATATTGAAGACAACTCAAACTCCGAGCTGGCTTTCCAGTGGCCCAAGCATCTATG 224
QY 584 GGTGACCTTCCACTATTAACACCGCAATATATGTTTTCAGAGCCATGCACTTTGAT 643
Db 225 GAATCCCATCTGTCATCAATTTCTGCCAATTTAGCTGTTATTTCTTGGCTTGGAGAAAGCT 284
QY 644 CGCAGCTAACCAAAAGAGCGCTTTGTTATCATATTTGATTACGATTTTCAACGAAGAT 703
Db 285 TAAACCTTGTATCACCAGATGAGTGAAGC-----TTTTTACCGCCAGC 329
QY 704 TGATCAATCTCATAGGGGACAAAGCTTTGGATATATATCTGGAGAGACTTTTCTGCTGAAA 763
Db 330 TTTTGGAACTCCATCAGGGACAAAGCTTAGATATTTACTGGAGGGA-----TAATTACA 383
QY 764 CATACCTACTCAGGAGATGTTTGAATATGTTTATGAATTAATAAAGAGCGGCTTTTCA 823
Db 384 CTTGTCCCACTGAAGAAGATATTAAGCTATGGTGTCTGCAGAAACAGGTGGACTGTTTG 443
QY 824 GATTAACGTTGAGACTCATGGAAGCGCTCTCTCTCTCTCACACCAACCGCCATTCGTTGG 883
Db 444 GATTAGAGTAGTCTCATGCGAGTTGTTCTCTGATTACAAAGAGATTTTAAACCGCTAC 503
QY 884 TTTCTTTTCAATAATCTTCTGGGTATTTATTTATCATGATTAGAGATGATTACTTGAATTTGA 943
Db 504 TT-----AATACACTTGGCTCTTTTCCAAATTAGGATGATTATGCTAATCTAC 554
QY 944 AAGATTTCCAAATGTCCAGCGAAAGAGCTTTGCTGAGGACATTACAGAGGGGAAGTTAT 1003
Db 555 ACTCCAAGATATATAGTGAACAAAGTTTGTGAGATCTGCAGAGGGGAAGTTCT 614
QY 1004 CTTTTCCTCATCGTCCAGCGCTT 1026
Db 615 CATTTCTCTACTATTCATGCTATT 637
```

Search completed: October 17, 2004, 03:24:28
Job time : 4884.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 17:43:59 ; Search time 128.441 Seconds
(without alignments)
8682.774 Million cell updates/sec

Title: US-10-041-018-1

Perfect score: 1569

Sequence: 1 aatattacatatagatag.....aatgcgggtaattcgga 1569

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1569	100.0	1569	2	US-08-761-344-3
2	1005	64.1	1005	2	US-08-761-344-1
3	114	7.3	903	1	US-08-469-665-1
4	114	7.3	903	2	US-09-038-596-1
5	114	7.3	903	5	PCT-US95-00421-1
6	70.6	4.5	1882	3	US-09-091-725-14
7	65	4.1	498	4	US-09-248-796A-3882
8	52.6	3.4	832	4	US-09-621-976-2813
9	49.8	3.2	1141	4	US-09-806-708B-22
10	48.2	3.1	1664976	4	US-08-916-421B-1
11	48.2	3.1	1664976	4	US-09-692-570-1
12	45.2	2.9	392000	4	US-10-027-983-11
13	44.8	2.9	486	4	US-09-248-796A-3883
14	44.4	2.8	6583	4	US-10-204-708-25
15	43	2.7	1055	4	US-09-806-708B-23
16	42.8	2.7	1141	4	US-09-806-708B-22
17	41.2	2.6	5152	4	US-10-204-708-73
18	41	2.6	765	4	US-09-601-198-160
19	40.8	2.6	1956	3	US-08-559-896B-1
20	40.8	2.6	1956	4	US-09-351-794A-1
21	40.6	2.6	6656	4	US-10-204-708-75
22	40	2.5	993	1	US-08-705-377-2
23	40	2.5	993	1	US-08-705-377-4
24	40	2.5	993	1	US-09-052-962-2
25	40	2.5	993	2	US-09-052-962-2
26	40	2.5	993	2	US-09-053-068-2
27	40	2.5	993	2	US-09-053-068-4

28 39.8 2.5 687 4 US-09-248-796A-8427 Sequence 8427, Ap
29 39 2.5 832 4 US-09-621-976-2813 Sequence 2813, Ap
30 39 2.5 6020 4 US-10-204-708-7 Sequence 7, Appli
31 39 2.5 17327 1 US-07-906-871-15 Sequence 15, Appl
32 38.8 2.5 3701 3 US-08-845-258-10 Sequence 10, Appl
33 38.8 2.5 3701 3 US-08-990-571-10 Sequence 10, Appl
34 38.8 2.5 3701 3 US-08-723-142A-10 Sequence 10, Appl
35 38.8 2.5 3701 4 US-09-528-784A-10 Sequence 10, Appl
36 38.8 2.5 3701 4 US-09-569-098A-10 Sequence 10, Appl
37 38.6 2.5 4770 3 US-09-000-094-45 Sequence 45, Appl
38 38.6 2.5 4770 4 US-10-011-749-45 Sequence 45, Appl
39 38.6 2.5 4770 4 US-09-000-084-45 Sequence 45, Appl
40 38.4 2.4 990 1 US-08-410-167A-1 Sequence 1, Appli
41 38.4 2.4 993 1 US-08-705-377-1 Sequence 3, Appli
42 38.4 2.4 993 1 US-08-705-377-6 Sequence 6, Appli
43 38.4 2.4 993 2 US-09-052-962-1 Sequence 1, Appli
44 38.4 2.4 993 2 US-09-052-962-3 Sequence 3, Appli
45 38.4 2.4 993 2 US-09-052-962-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-08-761-344-3

; Sequence 3, Application US/08761344

; Patent No. 5912154

; GENERAL INFORMATION:

; APPLICANT: Ferro-No. 5912154ick, Susan

; APPLICANT: Jiang, Yu

; TITLE OF INVENTION: GERANYLGERANYL DIPHOSPHATE SYNTHASE

; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sheridan Ross P.C.

; STREET: 1700 Lincoln Street

; CITY: Denver

; STATE: CO

; COUNTRY: U.S.A.

; ZIP: 80203

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/761,344

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Connell, Gary J.

; REGISTRATION NUMBER: 30,020

; REFERENCE/DOCKET NUMBER: 3161-14

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 303/863-9700

; TELEFAX: 303/862-0223

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1569 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-761-344-3

Query Match 100.0%; Score 1569; DB 2; Length 1569;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1569; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATATTACATAGATAGGACAGCCGCGATTTTTCATCTGAAAGGTAATCTTATT 60

Db 1 AATATTACATAGATAGGACAGCCGCGATTTTTCATCTGAAAGGTAATCTTATT 60

QY 61 ATTATAGTGGTATCCAAAGTTTCCGGCTTCAGCATACAGAAATTACGTGTTTTGCAT 120
DB 61 ATTATAGTGGTATCCAAAGTTTCCGGCTTCAGCATACAGAAATTACGTGTTTTGCAT 120
QY 121 ATGTTATCTGATCATGTTGCTTACTACATATTTTCTTGGCTTCGGCTTCTTT 180
DB 121 ATGTTATCTGATCATGTTGCTTACTACATATTTTCTTGGCTTCGGCTTCTTT 180
QY 181 GACGTTTTTTTTGAAGCAAAAAAAGTCAAGACAGATGCTTTACAAAACCAATGTAAGC 240
DB 181 GACGTTTTTTTTGAAGCAAAAAAAGTCAAGACAGATGCTTTACAAAACCAATGTAAGC 240
QY 241 TCAITTTTCAAGAACTTAAATGAGAGAGAAAGAGTTTACGAGTCTGAAAAATCA 300
DB 241 TCAITTTTCAAGAACTTAAATGAGAGAGAAAGAGTTTACGAGTCTGAAAAATCA 300
QY 301 ATGGAGGCCAAGATAGATGAGCTGATCAATATGATCCTGTTGGTCCAGCCAAAATGAA 360
DB 301 ATGGAGGCCAAGATAGATGAGCTGATCAATATGATCCTGTTGGTCCAGCCAAAATGAA 360
QY 361 AGCTTGATTTCAAAACCTTTATATCATCTCTTTTGAACCTGGCAAGAACTTTAGACTA 420
DB 361 AGCTTGATTTCAAAACCTTTATATCATCTCTTTTGAACCTGGCAAGAACTTTAGACTA 420
QY 421 AATTTAATAGTTTCAAAATTAACAGAGTTATGAAATTTGCCCAAGACAGCTGGCCATAGT 480
DB 421 AATTTAATAGTTTCAAAATTAACAGAGTTATGAAATTTGCCCAAGACAGCTGGCCATAGT 480
QY 481 TCGCAAAATTTGAGCTCTTGCATATATCCAGCTTTTAAATCGAGATATAGAAATAT 540
DB 481 TCGCAAAATTTGAGCTCTTGCATATATCCAGCTTTTAAATCGAGATATAGAAATAT 540
QY 541 GTCCTCTGAGAAAGGAGCAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTATA 600
DB 541 GTCCTCTGAGAAAGGAGCAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTATA 600
QY 601 AACACCGCAAAATTTATATGATTTTTCAGAGCATCACTTGTATCGCAGCTTAACCAAAA 660
DB 601 AACACCGCAAAATTTATATGATTTTTCAGAGCATCACTTGTATCGCAGCTTAACCAAAA 660
QY 661 GAGCCTTTGATCATATTTGATTTGATTTTCAAGCAAGAAATGATCAATCTACATAGG 720
DB 661 GAGCCTTTGATCATATTTGATTTGATTTTCAAGCAAGAAATGATCAATCTACATAGG 720
QY 721 GGAAGAGCTTGATATATCTGAGAGACTTTCTGCTGAATCACTACTACTCAGGAG 780
DB 721 GGAAGAGCTTGATATATCTGAGAGACTTTCTGCTGAATCACTACTACTCAGGAG 780
QY 781 ATGTTATTTGAATATGTTATGATTAATAACAGGGGCTTTTCAGATTAAGTTGAGACTC 840
DB 781 ATGTTATTTGAATATGTTATGATTAATAACAGGGGCTTTTCAGATTAAGTTGAGACTC 840
QY 841 ATGGAAGCGCTGTCTCTCTCTCACCACCGGCAATCGTTGGTTCCTTTCAATAATCTTT 900
DB 841 ATGGAAGCGCTGTCTCTCTCTCACCACCGGCAATCGTTGGTTCCTTTCAATAATCTTT 900
QY 901 CTGGGTATTTATTCAGATAGAGATGATTTGAAATTTGAAAGATTTCCAAATGTC 960
DB 901 CTGGGTATTTATTCAGATAGAGATGATTTGAAATTTGAAAGATTTCCAAATGTC 960
QY 961 AGCGAAAAAGCTTTTGTGAGGACATTAACAGAGGGGAGTTATCTTTCCCATCGTCCAC 1020
DB 961 AGCGAAAAAGCTTTTGTGAGGACATTAACAGAGGGGAGTTATCTTTCCCATCGTCCAC 1020
QY 1021 GCCCTTAACCTCAATAAAAGGTCAACTGAGCAACAATGAAATTTCTAAGAAAT 1080
DB 1021 GCCCTTAACCTCAATAAAAGGTCAACTGAGCAACAATGAAATTTCTAAGAAAT 1080
QY 1081 CTCCTGTTGAGGACAGTGTAAAGATATAAACTAAAGCTGATTCAATACTGGAATTC 1140
DB 1081 CTCCTGTTGAGGACAGTGTAAAGATATAAACTAAAGCTGATTCAATACTGGAATTC 1140
QY 1141 GACACCAATTCATTGGGCTCACCAAAAAATTTTAAATCAATAGTGAATATGATAAA 1200

DB 1141 GACACCAATTCATTGGGCTCACCAAAAAATTTTAAATCAATAGTGAATATGATAAA 1200
QY 1201 AATGATTAAGAAATAAGTATTTTACCTGATTGGCTTCGCAATTCGACACCGCCACCAAT 1260
DB 1201 AATGATTAAGAAATAAGTATTTTACCTGATTGGCTTCGCAATTCGACACCGCCACCAAT 1260
QY 1261 TTACATACGCAATTTGTTATATATATATAGACCACTTATCCGAATTTGTAATAAATTGATC 1320
DB 1261 TTACATACGCAATTTGTTATATATATATAGACCACTTATCCGAATTTGTAATAAATTGATC 1320
QY 1321 AATCAAAATTTAGTGGAGGAAGATAGTCAAGAAATAAGCCCTTCTCTCTCTCTCGCATC 1380
DB 1321 AATCAAAATTTAGTGGAGGAAGATAGTCAAGAAATAAGCCCTTCTCTCTCTCTCGCATC 1380
QY 1381 TATACATACGATTTTCATATATACGTTTTCATTTGCATCATCTTTTGTATATATCTCAAAAAGA 1440
DB 1381 TATACATACGATTTTCATATATACGTTTTCATTTGCATCATCTTTTGTATATATCTCAAAAAGA 1440
QY 1441 TCTCTTAGTTCGCAAAATAGTCAAAATCTTCAAAATTTATAGCCTTTATTTTCCAGAT 1500
DB 1441 TCTCTTAGTTCGCAAAATAGTCAAAATCTTCAAAATTTATAGCCTTTATTTTCCAGAT 1500
QY 1501 TTCTGAAATCTCTTTTATATATCAGCAGCGTTAATGCTAGCGGTTACTGTCAAAATCGCGGTA 1560
DB 1501 TTCTGAAATCTCTTTTATATATCAGCAGCGTTAATGCTAGCGGTTACTGTCAAAATCGCGGTA 1560
QY 1561 AATTCGCGA 1569
DB 1561 AATTCGCGA 1569

RESULT 2

US-08-761-344-1
; Sequence 1, Application US/08761344
; Patent No. 5912154
; GENERAL INFORMATION:
; APPLICANT: Ferro-No. 5912154ick, Susan
; APPLICANT: Jiang, Yu
; TITLE OF INVENTION: GERANYLGERANYL DIPHOSPHATE SYNTHASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/761,344
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J
; REGISTRATION NUMBER: 30,020
; REFERENCE/DOCKET NUMBER: 3161-14
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/863-9700
; TELEFAX: 303/862-0223
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1005 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:

```

; NAME/KEY: CDS
; LOCATION: 1..1005
US-08-761-344-1

```

Query Match	64.1%;	Score 1005;	DB 2;	Length 1005;
Best Local Similarity	100.0%;	Prod. No. 4.2e-253;		
Matches 1005;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	301	ATGAGGGCCAAGATAGATGAGCTGATCAATATGATCTGTGTTGGTCCAGCCAAATGAA	360	
Db	1	ATGAGGCCAAGATAGATGAGCTGATCAATATGATCTGTGTTGGTCCAGCCAAATGAA	60	
QY	361	AGCTTGATTTCAAACCTTAAATCAATCACTCTTTGAAACCTGCGAAGACTTTAGACTA	420	
Db	61	AGCTTGATTTCAAACCTTAAATCAATCACTCTTTGAAACCTGCGAAGACTTTAGACTA	120	
QY	421	AAATTAATAGTTCAAATTAACAGAGTTATGAATTTGGCCAAAGACCGCATAGTT	480	
Db	121	AAATTAATAGTTCAAATTAACAGAGTTATGAATTTGGCCAAAGACCGCATAGTT	180	
QY	481	TCGCAAAATTTGTGAGCTCTTCATATAATTCAGAGCCTTTTAAATCGACGATATAGAGATAAT	540	
Db	181	TCGCAAAATTTGTGAGCTCTTCATATAATTCAGAGCCTTTTAAATCGACGATATAGAGATAAT	240	
QY	541	GCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGGTGTACCTCCACTATA	600	
Db	241	GCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGGTGTACCTCCACTATA	300	
QY	601	AACACCGCAAAATATATGTAATTCAGAGCCATGCAACTTGTATCGCAGCTAACCAAAA	660	
Db	301	AACACCGCAAAATATATGTAATTCAGAGCCATGCAACTTGTATCGCAGCTAACCAAAA	360	
QY	661	GAGCCTTTGTATCATAAATTTGATTACCAATTTTCAAACGAAGATTTGATCAATCTACATAGG	720	
Db	361	GAGCCTTTGTATCATAAATTTGATTACCAATTTTCAAACGAAGATTTGATCAATCTACATAGG	420	
QY	721	GGCAAGGCTTGGATATATATCTCGAGAGACTTTCCTGCTGAAATCATCTACAGGAG	780	
Db	421	GGCAAGGCTTGGATATATATCTCGAGAGACTTTCCTGCTGAAATCATCTACAGGAG	480	
QY	781	ATGTAATTTGAATATGGTTATGAATAAACAGCGGCCCTTTTCAGATTAACGTTGAGACTC	840	
Db	481	ATGTAATTTGAATATGGTTATGAATAAACAGCGGCCCTTTTCAGATTAACGTTGAGACTC	540	
QY	841	ATGGAAGCGCTGTCTCCTCCTCACACCGGCCATTCGTTGGTTCCTTTCAATAATCTTT	900	
Db	541	ATGGAAGCGCTGTCTCCTCCTCACACCGGCCATTCGTTGGTTCCTTTCAATAATCTTT	600	
QY	901	CTGGGTATTTATTTACAGATTTAGAGATGATTAATGTAATTTGAAAGATTTCCAAATGTCC	960	
Db	601	CTGGGTATTTATTTACAGATTTAGAGATGATTAATGTAATTTGAAAGATTTCCAAATGTCC	660	
QY	961	AGCGAAAAAGGCTTTCCTCGAGACATTACAGAGGGGAAGTTATCTTTTCCATCTGCGCCAC	1020	
Db	661	AGCGAAAAAGGCTTTCCTCGAGACATTACAGAGGGGAAGTTATCTTTTCCATCTGCGCCAC	720	
QY	1021	GGCCTTAACTTCTACTAAAACGAAAGGTCAAACCTGAGCAACAATGAAATTTCAAGAAAT	1080	
Db	721	GGCCTTAACTTCTACTAAAACGAAAGGTCAAACCTGAGCAACAATGAAATTTCAAGAAAT	780	
QY	1081	CTCCTCTTGAGGACAAAGTGATAAAGATATAAACTAAAGCTGATTCAAATCTGGAATTC	1140	
Db	781	CTCCTCTTGAGGACAAAGTGATAAAGATATAAACTAAAGCTGATTCAAATCTGGAATTC	840	
QY	1141	GACCAAAATTCATTCGCCCTACACCAAAAAATTTTATTAATCAATTAGTGAATATGATAAAA	1200	
Db	841	GACCAAAATTCATTCGCCCTACACCAAAAAATTTTATTAATCAATTAGTGAATATGATAAAA	900	
QY	1201	AATGATAATCGAAATAAGTATTTACCTGATTTGGCTTCGCAATTCGACACCGCCCAAT	1260	
Db	901	AATGATAATCGAAATAAGTATTTACCTGATTTGGCTTCGCAATTCGACACCGCCCAAT	960	
QY	1261	TTACATGACGAATTTGTTATATATAATAGACCACTTATCCGAATTTG	1305	

Db 961 TTACATGACGAATTGTTATATATAATAGACCACTTATCCGAATTG 1005

RESULT 3

```

US-08-469-665-1
; Sequence 1, Application US/08469665
; Patent No. 5786193
; GENERAL INFORMATION:
; APPLICANT: GREENE, ET AL.
; TITLE OF INVENTION: Human Geranylgeranyl Pyrophosphate Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,665
; FILING DATE: June 6, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/00421
; FILING DATE: 11 JAN 95
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-377
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 903 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
US-08-469-665-1

```

Query Match	7.38; Score 114; DB 1; Length 903;
Best Local Similarity	52.1%; Pred. No. 3e-20;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;	
QY	359 AAGCTTGAATTCAAACCTTTATAATCACATCCTTTTGGAAACCTGGCAAGAACTTTAGAC 418
Db	
	26 AAGAATCTCTAGAACCCCTATAAATACATTACTTCAGTTACCAGGTAAACAAGTGAA 85
QY	419 TAAATTTAATAGTTCAAATTTAAACAGAGTTATGAATTTGCCCAAGACCGAGCTGGCCCATAG 478
Db	
	86 CCAAACTTTCACAGGCAATTTAACTATTGGCTGAAAGTTCCAGAGGACAAAGCTACAGATTA 145
QY	479 TTTTGC AAAATGTTGAGCTCTGCATAAATCCAGCCCTTTTAAATCGACGATATAGAAGATA 538
Db	
	146 TTATTGAAGTGACAGAAATGTTGCATAATGCCAGTTTACTCATCGATGATATTGAAGACA 205
QY	539 ATGCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAAATCTTCGGTGTACCTCCCATTA 598
Db	
	206 ACTCAAACCTCCAGCGTGGCTTTTCCAGTGCCCAACAGCATCTATGGAATCCCAATCTGCA 265
QY	599 TAAACACCGCAATATATGTTATTTTTCAGACCCATGCAACTTGTATCCGAGCTAACCCAA 658
Db	
	266 TCAATTCGCCAATACGUGTATTTCTCGCTTGGGAGAAAGTCT----- 310
QY	659 AAGAGCCTTTGTATCATATAATTTGATTACGATTTTCAACGAAGAATTTCAATCACTCATATA 718

Db 311 TAACCTTGATCACCAGATGAGTGAAGCTTTTACCGCAGCTTTTGAACCTCCATC 370
Qy 719 GGGGACAGGCTTGATATATACCTGAGAGACTTTCTGCTGAAATCATACCTACTCAGG 778
Db 371 AGGACAGGCTAGATATTTACTGAGGGA-----TAAATTACACTTGTGCCACTGAAG 424
Qy 779 AGATGATTTGAATGTTATGAAATAAAAGGGGGCTTTTTCAGATTAAGTTGAGAC 838
Db 425 AAGATATAAAGCTATGCTGCTGCAAGAAACAGGTGGAATGTTGGATTTAGCAGTAGGTC 484
Qy 839 TCATGAGAGCTGCTCTCTCTCTCACACCGCCATTTCTGTTGTTCTCTTTCATATAATC 898
Db 485 TCATGAGTGTCTCTGATTAACAAGAGATTTAAACCGCTACTT-----ATA 535
Qy 899 TTCTGGTATTTATTCAGATTTAGATGATTAATCTTGAATTTGAAAGATTTCCAAATGT 958
Db 536 CACTTGGCTCTTTTCCAAATAGGATGATTAATCTAATCTCACTCCAAAGATATA 595
Qy 959 CCAGCGAAAAGCTTTGCTGAGGACATTTACAGAGGGAGTATCTTTTCCCATCGTCC 1018
Db 596 GTGAAAACAAAAGTTGGGTGAGATCTGACAGAGGGAAGTTCTCATTTCTACTATTC 655
Qy 1019 ACGCCCTT 1026
Db 656 ATGCTATT 663

RESULT 4

US-09-038-596-1
; Sequence 1, Application US/09038596
; Patent No. 5928924
; GENERAL INFORMATION:
; APPLICANT: GREENE, ET AL.
; TITLE OF INVENTION: Human Geranylgeranyl Pyrophosphate Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA: US/09/038,596
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,665
; FILING DATE: June 6, 1995
; APPLICATION NUMBER: PCT/US95/00421
; FILING DATE: 11 JAN 95
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-377
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 903 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: cDNA
US-09-038-596-1

Query Match 7.3%; Score 114; DB 2; Length 903;
Best Local Similarity 52.1%; Pred. No. 3e-20;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;
Qy 359 AAAGCTTGATTTCAAAACCTTATATATCATCTCTTTTGAACCTTGGCAAGTATTAGAC 418
Db 26 AAAGATTTCTTTAGAACCTTATAATATCTTACTTCACTTACAGGTAAACAAAGTGAGAA 85
Qy 419 TAAATTTAATAGTTCAAAATTAACAGATTTATGAATTTGCCAAAGACACAGCTGGCCATAG 478
Db 86 CCAAACTTTTCACAGGCAATTAATCATGCTGGAAGTTCCAGAGACAGCTACAGATTATA 145
Qy 479 TTTCCAAATTTCTGAGCTCTTGCATAAATTCAGAGCTTTTAAATCGACGATATAGAAGATA 538
Db 146 TTAATGAGTGCACAGAAATGTTGCATAATGCCAGTTTACTCATGATGATATTTGAAGACA 205
Qy 539 ATGCTCCCTTGAGAAAGGGACAGACCACTTCTACATTAATCTTCGGTGTACCTCCACTA 598
Db 206 ACTCAAACTCCGAGCTGGCTTTCCAGTGGCCACAGCATCTATGGAATCCCATCTGTCA 265
Qy 599 TAAACACCGCAATTTATATATTTTTCAGAGCCATCACTTGTATCGCAGCTAAACACAAA 658
Db 266 TCAATTTCTGCCAATTAAGTGTATTTTCTTGGCTTGGAGAAAGTCT----- 310
Qy 659 AAGAGCTTTGTATCATTAATTTGATTACGATTTTCAACGAGAAATTTGATCAATCTACATA 718
Db 311 TAACTCTTGATCACCAGATGAGTGAAGCTTTTACCAGGCTTTTGGAACTCCATC 370
Qy 719 GGGGACAGGCTTGATATATATCTGAGAGACTTTTCTGCTGAAATCATACCTACTCAGG 778
Db 371 AGGACAGGCTTAGATATTTACTTGGAGGGA-----TAAATTACACTTGTCCCACTGAAG 424
Qy 779 AGATGATTTGAATGTTATGTTTGAATAAAACAGGGGGCTTTTTCAGATTAACGCTTGAGAC 838
Db 425 AAGATATAAAGCTATGTTGCTGACAGAAACAGGTGGAATTTTGGATTTAGCAGTAGGTC 484
Qy 839 TCATGGAAGCGCTGCTCTCTCTCACACCGCCATTTCTGTTGGTTCCTTTTCATTAATC 898
Db 485 TCATGCACTTGTCTCTGATTAACAAGAGATTTTAAACCGCTACTT-----ATA 535
Qy 899 TTCTGGTATTTATTCAGATTTAGATGATTTGATTTGAATTTGAAAGATTTTCCAAATGT 958
Db 536 CACTTGGCTCTTTTCCAAATAGGATGATTAATCTAATCTCACTCCAAAGATATA 595
Qy 1019 ACGCCCTT 1026
Db 656 ATGCTATT 663

RESULT 5

PCT-US95-00421-1
; Sequence 1, Application PC/TUS9500421
; GENERAL INFORMATION:
; APPLICANT: GREENE, ET AL.
; TITLE OF INVENTION: Human Geranylgeranyl Pyrophosphate Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00421
FILING DATE: Concurrently
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-257
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
PCT-US95-00421-1

Query Match
Best Local Similarity 72.1%; DB 5; Length 903;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;

359 AAAGCTGATTTCAAAACCTTATATACATCCCTTTGAAACCTGGCAAGACTTTAGAC 418
|||||
26 AAGAAATCTTCTAGAACCTTATAATCTTACTTTCAGTTACCAAGGTAAACAAGTGAA 85
419 TAAATTAATAGTTCAAATTAACAGAGTTATGAATTTGCCCAAGACCAAGCTGGCCATAG 478
86 CCAAACTTTCACAGCAATTAATCATTTGCTGAAAGTTCCAGAGGACAGCTACAGATTA 145
479 TTTTCGCAATTTGTTAGCTCTTGATATATTCAGCCCTTTTAAATGCAATATAGAGATA 538
146 TTATTGAAGTACAGAAATTTGATATATGCAATATGCAATATGCAATATGCAATATG 205
539 ATGCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTGGTGTACCTCCACTA 598
206 ACTCAAACTCCGAGTGGCTTCCAGTGGCCACAGCATCTATGGAATCCCATCTGTC 265
599 TAAACACCGCAATATATATGTTATTTACAGAGCCATGCACTTGTATCGCAGCTAACCA 658
266 TCAATCTGCAATTTACGTGTATTTCTTGGCTTGGAGAACTCT----- 310
659 AAGAGCTTTGTATCAATAATTTGATTTAGATTTTCAAGAGAAATGATCAATCTACATA 718
311 TAACCTTGTATCAACCCAGATGCAAGAGCTTTTACCCGCGAGCTTTTGGAACTCCATC 370
719 GGGGACAAAGCTTGGATATATCTACGAGAGACTTCTGCTGGAATCATACCTACTCAGG 778
371 AGGACAAAGCTTGGATATTTTACTTGGAGGA-----TAAATCACTTGTCCCACTGAAG 424
779 AGATGTTATTTGAATGCTTATGATTAATAAAGAGCGGCGCTTTTTCAGATTAACGTTGAG 838
425 AAGAATATAAAGCTATGCTGTCGAGAAACAGAGTGGACTTTTGGATTTAGCAGTAGTC 484
839 TCATGGAAGCGCTGCTCCCTTCTCAACAGCGGCCAATGCTGGTGGTTCCTTTTCAATAATC 898
485 TCATGCAATTTGTTCTCTGATTAACAAAGAGATTTAAACCGCTACTT-----AATA 535
899 TTCTGGGTATTTATTTATCAGATTTAGATGATTTGATTTGAAAGATTTTCCAAATGT 958
536 CACTTGGGCTCTTTTCCAAATTTAGGATGATTTGATTTGATTTTCCCAAGATATA 595
959 CCAGCGAAAAGGCTTTGCTGAGCAATTTACAGAGGGGAAATTTATCTTTTCCCATCTCC 1018
596 GTGAAAACAAAGTTTGGTGAAGATCTGACAGAGGGAAGTTCTCAATTTCTACTATTC 655
1019 ACSCCCTT 1026

Db 656 ATGCTATT 663

RESULT 6

US-09-091-725-14
Sequence 14, Application US/09091725
Patent No. 6329141
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Improved methods for transforming Phaffia
TITLE OF INVENTION: and recombinant DNA for use therein
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster llp
STREET: 2000 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: DC
COUNTRY: United States of America
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/091,725
FILING DATE: 23-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95203620.0
FILING DATE: 22-DEC-1995
APPLICATION NUMBER: EP 96200943.7
FILING DATE: 11-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: E. Victor Donahue
REGISTRATION NUMBER: 35,492
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1882 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Phaffia rhodozyma
FEATURE:
NAME/KEY: CDS
LOCATION: 82..1212
OTHER INFORMATION: /product= "PrortE"
US-09-091-725-14

Query Match 4.5%; Score 70.6; DB 3; Length 1882;
Best Local Similarity 51.8%; Pred. No. 9e-09;
Matches 160; Conservative 0; Mismatches 149; Indels 0; Gaps 0;
QY 352 CAAATGAAAGCTTGATTTCAAACCTTATAATCAATCTCTTTGAAACCTGGCAAGAC 411
Db 130 CAGGATGATCTGCTCTCTTGAACCTATCACTACTAGGAAGAACCTTGAAGAAGAA 189
QY 412 TTTAGACTAAATTAATAGTTCAAAITAAACAGAGTTATGAATTTGCCAAAGACCGCTG 471
Db 190 ATTTCGATCAACTCATCTCGAGGCTTTCACTATTTGGATCTCAAGAGGAGGATCTC 249
QY 472 GCCATAGTTTCGCAAAATTTGAGCTTTGCATAATTCAGCCCTTTTATTCGAGATATA 531
Db 250 GAGTCAATCAGAACGTTTGTGGCATGCTACATACCGCTAGCTTATTAATGGACGATG 309
QY 532 GAAGATAATGCTCCCTTTGAGAGGGGACAGACACTTCTCACTTAATCTTTCGGGTACCC 591
Db 310 GAGGATTCATCGTCTCTCAGCGTGGGTGCGCTGTGGCCATCTAATTTACGGGATTCGG 369


```
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
```

US-08-916-421B-1

```
Query Match 3.1%; Score 48.2; DB 4; Length 1664976;
Best Local Similarity 48.1%; Pred. No. 0.11;
Matches 137; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 368 TTTCAAAACCTTATATCATCATCTCTTTTGAACCTGGCAAGAACTTTAGACTAAATTTAA 427
Db 688622 TATATAAGCGCTCAAAACATCTTCTATTTGCTGGAGGAAGAGAAATTAGGCCATATTTAA 688681

QY 428 TAGTTCAATTAACAGAGCTTATGAATTTGCCAAAGACACAGCTGGCCCATAGTTTCGCAA 487
Db 688682 CTGTAGTAACCTTATATGTTGAAGAAAGAGATTTAGAGAGGTTTTCAGCCGCTGCTG 688741

QY 488 TTGTTGAGCTCTTGCATAAATCCAGCCTTTTAATCGACGATATAGAAATAGTCCCT 547
Db 688742 CAGTAGAGTTAATTCACAACTACACCTTAATACATGATGACATTATGGACAATGATGATG 688801

QY 548 TGAGAAGGGGACAGACCACTTCTCACTTAATCTTGGTGTACCTCCCACTATAAACACCG 607
Db 688802 AGAGGAGAGGAAACCAACAGTTTCATGTTGTTCTATGGAGAGCAATGGCTATCTTAGCTG 688861

QY 608 CAAATTATATGTTATTCAGAGCCATGCAACTTGTATCGCAGCTAA 652
Db 688862 GAGATTATATATGCTAAAGCTTTTGAAGCAGTTTCAAGATAA 688906

RESULT 11
US-09-692-570-1
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
; Patent No. 6797466
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275C1
; CURRENT APPLICATION NUMBER: US/09/692,570
; CURRENT FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; PRIOR APPLICATION NUMBER: US 08/916,421
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
```

```

; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (103398)..(103398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)

```

```

; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g

```

```

FEATURE:
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g

Query Match      3.1%; Score 48.2; DB 4; Length 1664976;
Best Local Similarity 48.1%; Pred. No. 0.11;
Matches 137; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 368 TTCTAAACCTTATATACATCATCTTTTGAACCTGGCAGAACCTTTAGACTAAATTTAA 427
Db 688622 TATATAACCGTCAAAACATCTTCTATTGCTGGAGGAAGAGAAATAGGCCATATTTAA 688681
QY 428 TAGTTCAAAATTAACAGAGTTATGAATTTGCCAAAGACAGCTGGCCATAGTTTCGCAA 487
Db 688682 CTGTAGTAACTTATATGTTGAAGAAAGACATATTTGAGGAGGTTTTCGCGCGCTGCTG 688741
QY 488 TTGTTGAGCTCTTGCAATAATCCAGCCTTTTAAATGACGATATAGAGATATAGTCTCCCT 547
Db 688742 CAGTAGAGTTAATTCACAACTACACCTTAATACATGATGACATATGGAACAATGATG 688801
QY 548 TGAGAGGGGACAGACCACTTCTACCTTAATCTCGGTGTAACCTCCACTATAAACACCG 607
Db 688802 AGAGGAGAGGAAACCAACAGTTTCATGTTGCTATGAGGAGGCAATGGCTATCTTAGCTG 688861
QY 608 CAAATATATATGTTATTTTCAGAGCCATGCACTTGTATCGCAGCTAA 652
Db 688862 GAGATTATATATGCTAAAGCTTTTGACGACGTTTCAAGATAA 688906

```

RESULT 12

```

US-10-027-983-11
Sequence 11, Application US/10027983
Patent No. 661762
GENERAL INFORMATION:
APPLICANT: Kenneth W. Dobie
APPLICANT: Mark P. Roach
TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
FILE REFERENCE: RTS-0340
CURRENT APPLICATION NUMBER: US/10/027,983
CURRENT FILING DATE: 2001-12-18
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 11
LENGTH: 392000
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: 137740
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 137742
OTHER INFORMATION: unknown
NAME/KEY: misc feature
LOCATION: (138122)..(138221)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: unsure
LOCATION: 14507
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 151967
OTHER INFORMATION: unknown
NAME/KEY: misc feature
LOCATION: (151967)..(1542066)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: unsure
LOCATION: 154217
OTHER INFORMATION: unknown
NAME/KEY: misc feature

```

```

LOCATION: (164037)..(164136)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (174657)..(174756)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (186224)..(186323)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (195242)..(195341)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: unsure
LOCATION: 202703
OTHER INFORMATION: unknown
NAME/KEY: misc feature
LOCATION: (202771)..(202870)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (206246)..(215602)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (218126)..(218225)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (220360)..(220459)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (222717)..(222816)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (223981)..(224080)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (227487)..(227586)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (230157)..(230256)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (232299)..(232398)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (236552)..(236651)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: (238789)..(248788)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: exon
LOCATION: (118288)..(119101)
OTHER INFORMATION: exon 1C
NAME/KEY: exon:intron junction
LOCATION: (151129)..(151130)
OTHER INFORMATION: exon 5:intron 5
NAME/KEY: exon:intron junction
LOCATION: (299248)..(299249)
OTHER INFORMATION: exon 9:intron 9
NAME/KEY: exon:intron junction
LOCATION: (348578)..(348579)
OTHER INFORMATION: exon 10:intron 10
NAME/KEY: intron
LOCATION: (348579)..(381838)
OTHER INFORMATION: intron 10
NAME/KEY: intron:exon junction
LOCATION: (386185)..(386186)
OTHER INFORMATION: intron 11:exon 12
US-10-027-983-11

```

Query Match 2.9%; Score 45.2; DB 4; Length 392000;
Best Local Similarity 50.0%; Pred. No. 0.36;
Matches 113; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 598 ATAAACCCCAATATATGCTATTTTCAGAGCCATGCACTTGTATCGCAGCTAACACA 657

```
Db 161142 ATAAATAACATCTCACTGTGTCATTCAGCTATATTCAGCATATGAGATCAGAAAGAAA 161201
QY 658 AAAGAGCCTTTTGTATCAVAATTTGATTACGATTTTCAACGAGAATTTGATCAATCTACAT 717
Db 161202 AAAAGAGCTAGTGTTTAAAGCAAGTTTATCTCCACAAAGAAATTAGAAATTTATAT 161261
QY 718 AGGGAGACAGGCTTGATATATATCTGAGAGAGCTTTCTGCTGCTGAAATCATACCTACTCAG 777
Db 161262 CTTGTAGAGAGTGATGAAATTTGCCATCTCTTCTTTCAGTCAAGAAAAAACCAGCAG 161321
QY 778 GAGATGATTTTGAATATGTTATGATTAATAACAGAGCGGCTTTTCA 823
Db 161322 TGCTGTATATGATGAAATTTTAACTAAGATAGTAATTAATTTTAA 161367

RESULT 13
US-09-248-796A-3883
; Sequence 3883, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3883
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-3883

Query Match 2.9%; Score 44.8; DB 4; Length 486;
Best Local Similarity 52.7%; Pred. No. 0.028;
Matches 97; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 469 CTGGCCATAGTTTCGCAATTTGTTGAGCTCTTGCAATATTCAGATTCACGCCCTTTTAATCGAGAT 528
Db 235 CTTGCAAGATAGTGAATTAATTTTCAGTTTTCATAACTCTTCATTTATGATAGATGAT 294
QY 529 ATAGAGATATGCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGA 588
Db 295 ATTGAAGATGATTCAAATATCGAGCGGGTCCAGTGGCACATGAAATATGGGATC 354
QY 589 CCCTCCACTATAACCGCAAAATATATGATTTATTCAGAGCCATGCAACTTGTATCGCAG 648
Db 355 CCAACACATAAATTTGGTGAATCTATGATTTTGTGGCGTTACACCAAGCTCAACAA 414
QY 649 CTAA 652
Db 415 TTAA 418

RESULT 14
US-10-204-708-25/c
; Sequence 25, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
```

```
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 25
; LENGTH: 6583
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-25

Query Match 2.8%; Score 44.4; DB 4; Length 6583;
Best Local Similarity 51.0%; Pred. No. 0.11;
Matches 105; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 1022 CCCTTAAGTTCACATAACGAAAGGTCAAACTGAGCAACACAAATGAAATCTTAAGAAATTC 1081
Db 5235 CCTCAACACCAATTTCAAAAAAACAATCTCTTACCTCTCAATAATATCTTAAAAAAAT 5176
QY 1082 TCCTGTTGAGGACAGTGATTAAGATATAAACTAAAGCTGAATCAATCTGGAATTCG 1141
Db 5175 TTCTCTCACTCCGAAAAACGAATATAATTAATTCATAAAATTTACATAAATTTCTTCT 5116
QY 1142 ACACCAATTCATTGGCTTACACCAAAAAATTTTATTAATCAATTAATGATGATAAAA 1201
Db 5115 AAACCAACAAATATCTCGAATAATTAATTTCTTCTTAACTGTAATAAAATACTAAAT 5056
QY 1202 ATGATAATGAAATAAGTATTTACCT 1227
Db 5055 ACGATAATAAAAAATACTATATACAT 5030

RESULT 15
US-09-806-708B-23/c
; Sequence 23, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 23
; LENGTH: 1055
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1055)
; OTHER INFORMATION: consensus sequence of A.t. and L.a. FAE1 promoters
US-09-806-708B-23

Query Match 2.7%; Score 43; DB 4; Length 1055;
Best Local Similarity 21.2%; Pred. No. 0.12;
Matches 126; Conservative 157; Mismatches 299; Indels 12; Gaps 2;

QY 886 CCCTTCATAAATCTCTGGGTATATATTCAGATGAGATGATTAATTTGAAATTTGAA 945
Db 634 CATNTWARATNNNTTTRAGGAATNTKGTSTTAGWMTGATTTGGNNNNNNNNNNNNNNNN 575
QY 946 GATTTCCAAATGTCAGCGAAAAAGGCTTTGCTGAGGACATTAACAGAGGGGAGTATCT 1005
Db 574 CTCTGACTAATATGATGATSAKURKATWTCTYRRRWANGWRTWTANWRCARWRTYTRY 515
```

```
QY 1006 TTTCCTCCTGTCAGCCCTTAACTTCACTAAACGAAGTCAACTCAGCAACACAAT 1065
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
514 CKSTAMSCITYWGNAMWYAAYSAGNTSSWARWTTANNTAAGGYMRAAGTWAAWANNNN 455
QY 1066 GAAATTCCTAAGAAATTCCTCTG-----TTGAGGACAAAGTGATAAAGATATAAACTAAAAGC 1120
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
454 TRTRYAWWTRWARVTTTTYRSACNAKSNWRGATWRATMTTAAANNKAGAMWMTTAAGN 395
QY 1121 TGATTCAAAATCTGGAATTCGACACCAATTCATTGGCCCTACACCCAAAATTTTATTAATC 1180
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
394 NTMTAAAATKMWAAWACARAYCCNNNWAACAMKMKWAWTWKYAWGAACNNNNKTAMYC 335
QY 1181 AATTAGTGAATATGATAAATAAATGATAATGAAAATAAGTATTACCTGATTTGGCTTCGC 1240
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
334 RRAWMYSAWTTTWAA-----ATSWMKWYTTTTRKTTMAAAANNNNNNNNAKCKTTSA 282
QY 1241 ATTCCGACACCCGCCCAATTTACATGACGAATTTGTATATATATAATAGACCACTTATCCG 1300
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
281 MWAMWMAWCTCGARTWGGATYMAAACCTAAGAGNSMTYCWAAATYMGTTNNNNNNK 222
QY 1301 AATTGTGAAATAAATTTGATCAATCAAAATPAGTGGAGGAAGATAGTCAGAAAATAAGCCTT 1360
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
221 AWTTRKTMAWCAYTWTARWYKRYTCYATVYVTSMTYMGKMWYWARAAYYASTNCTW 162
QY 1361 CTCTCCTCCTCTTTTCGCATCTATACATAGATTTTCATATATATACGTTTCATTGCATCATCT 1420
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
161 STCRWKATARGWYAMRYYTWAKAGCNNNAKWRCAWYACATNNNAWAWYACAYWYTT 102
QY 1421 TTTGTATATATCTCAAAAAGATCTCTTAGTTTCGAAATAGTCAAAATCTTCAAAAT 1474
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
101 AGKAATNNKIAGKWKWYAMMKTTWYAWYTWCAACWMAKERTAAKWCWYTT 48
```

Search completed: October 17, 2004, 03:30:33
Job time : 132.441 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2004, 16:35:39 ; Search time 733.231 Seconds
(without alignments)
11232.937 Million cell updates/sec

Title: US-10-041-018-1

Perfect score: 1569

Sequence: 1 aatattacatatagatag.....aatcgccggttaattcgga 1569

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134896 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_23Sep04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001as:*

5: Geneseq2001bs:*

6: Geneseq2002as:*

7: Geneseq2002bs:*

8: Geneseq2003as:*

9: Geneseq2003bs:*

10: Geneseq2003cs:*

11: Geneseq2003ds:*

12: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1569	100.0	1569	12	Adm98581 Geranylge
2	1008	64.2	1008	6	Aal40786 DNA relat
3	1008	64.2	1008	6	Abk96782 S. cerevi
4	1008	64.2	1008	10	Adc49385 Yeast BTS
5	1008	64.2	1008	12	Adl14810 Yeast ger
6	568.6	36.2	599	10	Adk52980 Plant DNA
7	183.4	11.7	963	10	Adf75080
8	114	7.3	903	2	Aat31846 Human ger
9	114	7.3	903	2	Aav41555 Human ger
10	114	7.3	903	2	Aax86773 Human ger
11	114	7.3	1395	12	Adj10255 Human ger
12	114	7.3	1547	3	Aaz60889 cDNA enco
13	114	7.3	1692	5	Aaz60890 cDNA enco
14	114	7.3	2534	12	Adl63550 Human ova
15	114	7.3	2534	12	Adm98583 Geranylge
16	113.6	7.2	2529	3	Aaf18152 Lung canc
17	99.6	6.3	1032	8	Abz77169 Phoma bet
18	99.6	6.3	1888	5	Abv25136 Human pro
19	87.6	5.6	17131	3	Aaz60888 Human enco
20	87.6	5.6	51001	12	Adj10262 Human ger
21	84.2	5.4	2210	4	AbL08261 Drosophil

22	84.2	5.4	2222	12	Adm98601
23	74.8	4.8	630	10	ADD34746
24	70.6	4.5	1882	2	AAT72942
25	68.8	4.4	554	6	ABQ57629
26	68.4	4.4	716	6	ABQ58229
27	65.6	4.2	593	6	ABQ57793
28	62.4	4.0	1843	12	Adm98585
29	60	3.8	399	5	ABV49901
30	58.6	3.7	1095	8	ACA52518
31	56.4	3.6	6115	6	ABL33801
32	49.2	3.1	6802	6	ABL33231
33	48.8	3.1	1044	8	ACA23035
34	48.8	3.1	110000	2	AXA20248_05
35	48.8	3.1	111309	2	AXA20250_06
36	48.2	3.1	110000	2	AAV21209_06
37	48	3.1	2000	8	ADA71938
38	47.8	3.0	6127	6	ABL33614
39	47.8	3.0	2703	6	ABN67916
40	47.8	3.0	110000	6	ABN71527_13
41	47	3.0	383	5	ADL44116
42	46.4	3.0	6681	6	ABL32155
43	46.4	3.0	6681	6	ABL54304
44	46.4	3.0	10856	6	ABL32457
45	46.4	3.0	18133	6	ABK40018

ALIGNMENTS

RESULT 1

Adm98581

ID Adm98581 standard; DNA; 1569 BP.

XX Adm98581;

AC Adm98581;

DT 01-JUL-2004 (first entry)

XX Geranylgeranyl pyrophosphate synthase DNA #1.

DE Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;

XX Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;

KW diterpene precursor; diterpene synthase; defence toxin;

KW volatile defensive signal; pollinator attractant; photoprotectant.

XX Saccharomyces cerevisiae.

OS US2004072323-A1.

XX 15-APR-2004.

PD 07-JAN-2002; 2002US-00041018.

PF 05-JAN-2001; 2001US-02599890P.

XX (MATS/) MATSUDA S P T.

XX (HART/) HART E A.

PI Matsuda SPT, Hart EA;

XX WPI; 2004-373921/35.

DR New unicellular organisms comprising exogenous nucleic acids encoding a

XX geranylgeranyl pyrophosphate and a diterpene synthase, useful for

XX producing diterpenes and diterpene precursors.

PS Claim 2; SEQ ID NO 1; 38pp; English.

XX The invention relates to a unicellular organism for producing a diterpene

XX or diterpene precursor comprising an exogenous nucleic acid sequence

XX encoding a geranylgeranyl pyrophosphate synthase under the control of a

XX promoter operable in the organism, and an exogenous nucleic acid sequence

XX encoding a diterpene synthase under the control of a promoter operable in

XX the organism. The invention also relates to methods of producing a

XX diterpene or diterpene precursor and a method of isolating a diterpene

CC synthase comprising growing several cells in the presence of a
 CC polyaromatic resin to make a cell/resin mixture, where at least one of
 CC the cells further comprises at least one isolated and purified nucleic
 CC acid sequence of a yeast expression library, and the expression of the
 CC nucleic acid sequence is regulated by an inducible promoter under
 CC conditions where the expression is induced, filtering the cell/resin
 CC mixture, extracting the cell/resin mixture with alcohol to produce an
 CC organic eluent and analysing the organic eluent by a screening method
 CC including chromatography and/or spectroscopy, to identify the nucleic
 CC acid sequence encoding the diterpene synthase. The unicellular
 CC microorganism is useful as a diterpene or diterpene precursor producing
 CC system. Diterpenes, in plants, serve as defence toxins, volatile
 CC defensive signals, pollinator attractants and photoprotectants. This
 CC sequence represents DNA encoding a geranylgeranyl pyrophosphate synthase
 CC polypeptide of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 1569 BP; 516 A; 307 C; 260 G; 486 T; 0 U; 0 Other;

Query Match 100.0%; Score 1569; DB 12; Length 1569;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1569; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATATTACATATAGATATAGACCAAGCCGCGATTTTTCATCTGAAAGGTAACCTCTATT 60
 DB 1 AATATTACATATAGATATAGACCAAGCCGCGATTTTTCATCTGAAAGGTAACCTCTATT 60

QY 61 ATTATAGTGATATCAAGCTTCAACGCTTCCAGCATATAGAGAGAGAAATTAAGTGTGAT 120
 DB 61 ATTATAGTGATATCAACGCTTCCAGCATATAGAGAGAGAAATTAAGTGTGAT 120

QY 121 ATGTTATGCTGATCATGTTGATCTTACTACCAATTTTCTGCTGCGCTTCTCTTTT 180
 DB 121 ATGTTATGCTGATCATGTTGATCTTACTACCAATTTTCTGCTGCGCTTCTCTTTT 180

QY 181 GACGTTTTTTTGAAGCAAAAAAGTCAAGCAGATGCTTTACAAAAACCATGTAAGGC 240
 DB 181 GACGTTTTTTTGAAGCAAAAAAGTCAAGCAGATGCTTTACAAAAACCATGTAAGGC 240

QY 241 TCATTTTCAAGAGCTACTATATAGAAAGAGAAAGAGTTTACAGTCTGGAAATCA 300
 DB 241 TCATTTTCAAGAGCTACTATATAGAAAGAGAAAGAGTTTACAGTCTGGAAATCA 300

QY 301 ATGAGGCCAAGATAGATGAGCTGATCAATATGATCTTGTGTCAGCCCAAAATGAA 360
 DB 301 ATGAGGCCAAGATAGATGAGCTGATCAATATGATCTTGTGTCAGCCCAAAATGAA 360

QY 361 AGCTTGATTTCAAAACCTTTATATCAATCCCTTTTGAACCTTGGCAAGAACTTTAGACTA 420
 DB 361 AGCTTGATTTCAAAACCTTTATATCAATCCCTTTTGAACCTTGGCAAGAACTTTAGACTA 420

QY 421 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGAACAGCTGGCCATAGTT 480
 DB 421 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGAACAGCTGGCCATAGTT 480

QY 481 TCGCAATTTGTCAGCTTTGATATATCCAGCTTTTATATCGACGATATAGAGATAAT 540
 DB 481 TCGCAATTTGTCAGCTTTGATATATCCAGCTTTTATATCGACGATATAGAGATAAT 540

QY 541 GCTCCCTTGAGAGGGGAGCAGACCACTTCTCACTTAATCTTGGTACCCCTCCACTATA 600
 DB 541 GCTCCCTTGAGAGGGGAGCAGACCACTTCTCACTTAATCTTGGTACCCCTCCACTATA 600

QY 601 AACACCGCAATTTATATGTTTTCAGAGCCATGCAACTTGTATCGAGCTTAACCAAAA 660
 DB 601 AACACCGCAATTTATATGTTTTCAGAGCCATGCAACTTGTATCGAGCTTAACCAAAA 660

QY 661 GAGCCTTTGATCATATTTGATGATTTTCAACGAGAAATTTGATCAATCTACATAGG 720
 DB 661 GAGCCTTTGATCATATTTGATGATTTTCAACGAGAAATTTGATCAATCTACATAGG 720

QY 721 GGAACAAGGCTTGGATATATATCTGGAGAGACTTTCTGCGTGAATATACCTACTCAGGAG 780

DB 721 GGACAAGGCTTGGATATATATCTGGAGAGACTTTCTGCGTGAATATCATCTACTCAGGAG 780
 QY 781 ATGTTATTTGAATATGTTTATGAATAAACAAGCGGCGCTTTTCAGATTAACGTTGAGACTC 840
 DB 781 ATGTTATTTGAATATGTTTATGAATAAACAAGCGGCGCTTTTCAGATTAACGTTGAGACTC 840
 QY 841 ATGGAAGCGCTGCTCTCTCTCTCAACACCGGCCATTTGTTGGTTCTCTTTTCAATAATCTT 900
 DB 841 ATGGAAGCGCTGCTCTCTCTCTCAACACCGGCCATTTGTTGGTTCTCTTTTCAATAATCTT 900
 QY 901 CTGGGTTATTTATTCAGATATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
 DB 901 CTGGGTTATTTATTCAGATATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
 QY 961 AGCGAAAAGGCTTTGCTGAGACATTAACAGAGGGGAGTATCTTTTCCCATCTGTCAC 1020
 DB 961 AGCGAAAAGGCTTTGCTGAGACATTAACAGAGGGGAGTATCTTTTCCCATCTGTCAC 1020

QY 1021 GCCCTTAACTTCACTAAAAACGAAAGGTCAAACTGAGCAACACAAATCAAAATTTCAAGAA 1080
 DB 1021 GCCCTTAACTTCACTAAAAACGAAAGGTCAAACTGAGCAACACAAATCAAAATTTCAAGAA 1080

QY 1081 CTCTGTTGAGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
 DB 1081 CTCTGTTGAGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140

QY 1141 GACCAATTCATTTGCTGAGCAACCAAAATTTTATTAATCAATTAATGATGATGATGATGAT 1200
 DB 1141 GACCAATTCATTTGCTGAGCAACCAAAATTTTATTAATCAATTAATGATGATGATGAT 1200

QY 1201 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
 DB 1201 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260

QY 1261 TTACATGAGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
 DB 1261 TTACATGAGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320

QY 1321 AATCAAAATTTAGTGGAGGAGATAGTACAGAAATAAAGCCCTTCTCTCTCTCTCTCTCTCT 1380
 DB 1321 AATCAAAATTTAGTGGAGGAGATAGTACAGAAATAAAGCCCTTCTCTCTCTCTCTCTCT 1380

QY 1381 TATACATACGATTTTAT 1440
 DB 1381 TATACATACGATTTTAT 1440

QY 1441 TCTCTTAGTTCCGAAATAGTCAAAATTTTCAAAATTTATAGCCCTTTATATTTTTCACGAT 1500
 DB 1441 TCTCTTAGTTCCGAAATAGTCAAAATTTTCAAAATTTATAGCCCTTTATATTTTTCACGAT 1500

QY 1501 TTCTGAAACTCTTTTATATAGCAGCAGCTTAAATGCTAGCGGTTACTCTCAAAATCGCCGGA 1560
 DB 1501 TTCTGAAACTCTTTTATATAGCAGCAGCTTAAATGCTAGCGGTTACTCTCAAAATCGCCGGA 1560

QY 1561 AATTCGCGA 1569
 DB 1561 AATTCGCGA 1569

RESULT 2

AAL40786

ID AAL40786 standard; DNA; 1008 BP.

XX

AC AAL40786;

XX

DT 03-OCT-2002 (first entry)

XX

DE DNA relating to the production of prenol alcohol SEQ ID No 5.

XX

KW Prenol alcohol; mutated cell; squalene synthase gene; geometric isomer;

XX

KW industrial synthesis; isoprenoid-terpenoid compound; gene; ds.

XX

Saccharomyces cerevisiae.

WO200253747-A1.

11-JUL-2002.

20-DEC-2001; 2001WO-JP011215.

28-DEC-2000; 2000JP-00401701.

28-DEC-2000; 2000JP-00403067.

18-SEP-2001; 2001JP-00282978.

(TOYT) TOYOTA JIDOSHA KK.

Ohto C, Obata S;

WPI; 2002-548086/58.

P-PSDB; AA022195.

Production of prenol alcohols by culturing translationally-active mutated cells with reduced squalene synthase gene to express less transcriptional product.

Disclosure; Page 143-146; 262pp; Japanese.

The invention relates to a method for producing prenol alcohol comprising culturing mutated cells, having been mutated so as to reduce the amount of transcriptional product of squalene synthase gene transcriptional activity, and then collecting prenol alcohol from the culture medium. The method is for the production of prenol alcohols, which is for use in industrial synthesis of isoprenoid-terpenoid compounds particularly physiologically-active prenol alcohol geometric isomers. This polynucleotide sequence represents a DNA sequence encoding a protein relating to the method for producing prenol alcohol comprising culturing mutated cells

Sequence 1008 BP; 337 A; 139 C; 176 G; 296 T; 0 U; 0 Other;

Query Match 64.2%; Score 1008; DB 6; Length 1008;

Best Local Similarity 100.0%; Pred. No. 9.4e-229;

Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OS	XX	301	ATGAGGCCAAGATAGATGAGTCAATATGATCCTGTTGGTCCAGCCAAATGAA	360
PN	XX	1	ATGAGGCCAAGATAGATGAGTCAATATGATCCTGTTGGTCCAGCCAAATGAA	60
PD	XX	361	AGCTTGATTTCAAACTTATATACATCCTTTTGAAACCTGGCAAGACTTTAGACTA	420
XX	XX	61	AGCTTGATTTCAAACTTATATACATCCTTTTGAAACCTGGCAAGACTTTAGACTA	120
XX	XX	421	AATTTAATAGTTCAAAATTAACAGAGTTATGAAATTTGCCAAGACCAGCTGGCCATAGT	480
XX	XX	121	AATTTAATAGTTCAAAATTAACAGAGTTATGAAATTTGCCAAGACCAGCTGGCCATAGT	180
XX	XX	481	TGCAAAATTTGAGCTCTTTCATATATCCAGCCTTTTAAATCGAGATATAGAGATAAT	540
XX	XX	181	TGCAAAATTTGAGCTCTTTCATATATCCAGCCTTTTAAATCGAGATATAGAGATAAT	240
XX	XX	541	GCTCCCTTGAGAGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTATA	600
XX	XX	241	GCTCCCTTGAGAGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTATA	300
XX	XX	601	AACACCGCAATATATATCTTTTTCAGAGCCATGCACTTGTATCGAGCTAACCAAAA	660
XX	XX	301	AACACCGCAATATATATCTTTTTCAGAGCCATGCACTTGTATCGAGCTAACCAAAA	360
XX	XX	661	GAGCCTTTGTATCATATATTTGATTTTCAACGAAGAAATGATCAATCTCATAGG	720
XX	XX	361	GAGCCTTTGTATCATATATTTGATTTTCAACGAAGAAATGATCAATCTCATAGG	420
XX	XX	721	GGACAGGCTTGGATATATATCTGGAGAGACTTTCTGCTGAAATCATACCTACTCAGG	780
XX	XX	421	GGACAGGCTTGGATATATATCTGGAGAGACTTTCTGCTGAAATCATACCTACTCAGG	480

QY	781	ATGATATTTGAATATCGTTTATGAAATAAAACAGCGCGCTTTTTCAGATTACGTTGAGACTC	840
DB	481	ATGATATTTGAATATCGTTTATGAAATAAAACAGCGCGCTTTTTCAGATTACGTTGAGACTC	540
QY	841	ATGGAAGCGCTGCTCTCCCTTCTCACACACCGCCATTCGTTGGTTCCTTTCATAAATCTT	900
DB	541	ATGGAAGCGCTGCTCTCCCTTCTCACACACCGCCATTCGTTGGTTCCTTTCATAAATCTT	600
QY	901	CTGGGTATTTATTTATCAGATTAGAGATGATTACTTTGAAATTTGAAAGATTTCCAAATGTC	960
DB	601	CTGGGTATTTATTTATCAGATTAGAGATGATTACTTTGAAATTTGAAAGATTTCCAAATGTC	660
QY	961	AGCGAAAAGGCTTTGCTGAGACATATACAGAGGGGAAATTAATCTTTCCCATCTCCAC	1020
DB	661	AGCGAAAAGGCTTTGCTGAGACATATACAGAGGGGAAATTAATCTTTCCCATCTCCAC	720
QY	1021	GCCTTTAACTTCACTAAACGAAAGTCAAACTGAGCAACACAAATGAAATCTTAAAGATT	1080
DB	721	GCCTTTAACTTCACTAAACGAAAGTCAAACTGAGCAACACAAATGAAATCTTAAAGATT	780
QY	1081	CTCCTTTGAGGACAAAGTGAATAAAGATATAAACTTAAAGCTGATTCAATATCTGAAATTC	1140
DB	781	CTCCTTTGAGGACAAAGTGAATAAAGATATAAACTTAAAGCTGATTCAATATCTGAAATTC	840
QY	1141	GACACCAATTCATTGGCTTACACCAAAATTTTATTAATCAATAGTAGTAATGATAAAA	1200
DB	841	GACACCAATTCATTGGCTTACACCAAAATTTTATTAATCAATAGTAGTAATGATAAAA	900
QY	1201	AATGATATGAAATAAGTATTTACCTGATTTGGCTTCGCATTCGACACCGCCACCAAT	1260
DB	901	AATGATATGAAATAAGTATTTACCTGATTTGGCTTCGCATTCGACACCGCCACCAAT	960
QY	1261	TTACATGACGAATTTGTTATATATATATAGACCACTTATCCGAATTTGTA	1308
DB	961	TTACATGACGAATTTGTTATATATATATAGACCACTTATCCGAATTTGTA	1008

RESULT 3

ABK96782

ID ABK96782 standard; DNA; 1008 BP.

AC ABK96782;

DT 24-SEP-2002 (first entry)

XX S. cerevisiae prenol diphosphate synthase gene #2.

XX Prenol alcohol; prenol diphosphate synthase; geranylgeraniol;
XX hydroxymethylglutaryl-CoA reductase; farnesylgeraniol; gene; ss;
XX isopentenyl diphosphate delta-isomerase; mevalonate kinase;
XX mevalonate CoA acetyltransferase; isoprenoid-terpenoid compound.

OS Saccharomyces cerevisiae.

XX WO200253746-A1.

PN 11-JUL-2002.

XX 20-DEC-2001; 2001WO-JP011214.

XX 28-DEC-2000; 2000JP-00403067.

XX (TOYT) TOYOTA JIDOSHA KK.

XX Ohto C, Obata S, Muramatsu M, Nishi K, Totsuka K;

XX WPI; 2002-537944/57.

XX Production of prenol alcohols by culturing a transformant transferred
XX with e.g. prenol diphosphate synthase gene, for use in industrial
XX synthesis of e.g. physiologically-active isoprenoid-terpenoid compounds.

PS Example 2; Page 182-185; 335pp; Japanese.

XX The invention relates to a process for producing a prenyl alcohol comprising: (A) construction of a recombinant by transferring an expression recombinant DNA or a DNA for genome integration into a host which contains prenyl diphosphate synthase gene or its variant; and (B) collecting product from the culture medium. Also described is: (1) a method for producing prenyl alcohol in which the expression recombinant DNA contains: (a) a hydroxymethylglutaryl-CoA reductase gene or its variant; or (b) an isopentenyl diphosphate delta-isomerase gene; (2) a process for producing geranylgeraniol, comprising: (a) constructing a recombinant by transferring an expression recombinant DNA or a DNA for genome integration into a host which contains hydroxymethylglutaryl-CoA reductase gene or its variant; and (b) isolating the product; (3) a process for producing farnesylgeraniol in which the expression recombinant DNA also contains an isopentenyl diphosphate delta-isomerase gene, mevalonate CoA acetyltransferase gene, hydroxymethylglutaryl-CoA synthase gene, mevalonate kinase gene, or mevalonate diphosphate decarboxylase gene. The methods are used for the production of prenyl alcohols, particularly for use in industrial synthesis of isoprenoid-terpenoid compounds e.g. physiologically-active prenyl alcohols including geometric isomers. ABK96780-ABK96897 represent prenyl diphosphate synthase genes and related PCR primers used in the methods of the invention

SQ Sequence 1008 BP; 337 A; 199 C; 176 G; 296 T; 0 U; 0 Other;

Query Match 64.2%; Score 1008; DB 6; Length 1008;
 Best Local Similarity 100.0%; Pred. No. 9.4e-229;
 Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 ATGGAGGCCAAGATAGATGAGCTGATCAATATGATCCCTGTTGGTCCAGCCAAAATGAA 360
 DB 1 ATGGAGGCCAAGATAGATGAGCTGATCAATATGATCCCTGTTGGTCCAGCCAAAATGAA 60

QY 361 AGCTTGATTTCAAAACCTTATATATCATCATCTTTTGAACCTGCGCAAGAACTTTAGACTA 420
 DB 61 AGCTTGATTTCAAAACCTTATATATCATCATCTTTTGAACCTGCGCAAGAACTTTAGACTA 120

QY 421 AATTAAATGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGAACACGCTGGCCATAGTT 480
 DB 121 AATTAAATGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGAACACGCTGGCCATAGTT 180

QY 481 TCGCAATTTGTTGAGCTCTTGCAATATCCAGCTTTTAAATCGACGATATAGAAAGATAAT 540
 DB 181 TCGCAATTTGTTGAGCTCTTGCAATATCCAGCTTTTAAATCGACGATATAGAAAGATAAT 240

QY 541 GCTCCCTTGAGAAAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTATA 600
 DB 241 GCTCCCTTGAGAAAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTATA 300

QY 601 AACACGCGAAATATATGATTTTACAGAGCATCACTGTTATCGAGCTAACACACAAA 660
 DB 301 AACACGCGAAATATATGATTTTACAGAGCATCACTGTTATCGAGCTAACACACAAA 360

QY 661 GAGCCTTTGATCATATTTGATTTGATTTTCAAGAGAAATGATCAATCTACATAGG 720
 DB 361 GAGCCTTTGATCATATTTGATTTGATTTTCAAGAGAAATGATCAATCTACATAGG 420

QY 721 GGACAAAGCTTGGATATATCTGAGAGACTTTCTGCTCGAAATCATACCTACTCAGGAG 780
 DB 421 GGACAAAGCTTGGATATATCTGAGAGACTTTCTGCTCGAAATCATACCTACTCAGGAG 480

QY 781 ATGTATTTGAATGTTGATGATTAACACAGCGGGCTTTTCAGATTAACGTTGAGACTC 840
 DB 481 ATGTATTTGAATGTTGATGATTAACACAGCGGGCTTTTCAGATTAACGTTGAGACTC 540

QY 841 ATGGAAGCGTGTCTCTCTCTCAACACAGCGCAATTCGTTGGTTCCTTTCAATAAATCTT 900
 DB 541 ATGGAAGCGTGTCTCTCTCTCAACACAGCGCAATTCGTTGGTTCCTTTCAATAAATCTT 600

QY 901 CTGGGTATTTATTCAGATAGATGATTAATCTTGAATTTGAAAGATTTCCAAATGTCC 960

DB 601 CTGGGTATTTATTTATCAGATTAGAGATTACTTGAATTTGAAAGATTTCCAAATGTCC 660

QY 961 AGCGAAAAGGCTTTGCTGAGGACATTACAGAGGGGAAGTTATCTTTTCCCATCGTCCAC 1020

DB 661 AGCGAAAAGGCTTTGCTGAGGACATTACAGAGGGGAAGTTATCTTTTCCCATCGTCCAC 720

QY 1021 GCCCTTAACCTTCACTAAAACGAAAGGTCAAACCTGAGCAACACAATGAAATTTAAGAAAT 1080

DB 721 GCCCTTAACCTTCACTAAAACGAAAGGTCAAACCTGAGCAACACAATGAAATTTAAGAAAT 780

QY 1081 CTCTGTTGAGCAAGTATGATATAAAGTATAAAGTATAAAGTATAAAGTATAAAGTATAAAGT 1140

DB 781 CTCTGTTGAGCAAGTATGATATAAAGTATAAAGTATAAAGTATAAAGTATAAAGTATAAAGT 840

QY 1141 GACACCAATTCATGCGCTTACACCAAAATTTTATTAATCAATTAAGTATAAAGTATAAAGT 1200

DB 841 GACACCAATTCATGCGCTTACACCAAAATTTTATTAATCAATTAAGTATAAAGTATAAAGT 900

QY 1201 AATGATAATGAAATTAAGTATTTTACCTGATTTGGCTTCGATTCGACACCGCCCAAT 1260

DB 901 AATGATAATGAAATTAAGTATTTTACCTGATTTGGCTTCGATTCGACACCGCCCAAT 960

QY 1261 TTACATGACGAATTTGTTATATATATATATAGACCACTTATCCGAATTTGTA 1308

DB 961 TTACATGACGAATTTGTTATATATATATATAGACCACTTATCCGAATTTGTA 1008

RESULT 4
 ADC49385
 ID ADC49385 standard; DNA; 1008 BP.
 XX
 AC ADC49385;
 DT 18-DEC-2003 (first entry)
 XX
 DE Yeast BTS1 cDNA.
 XX
 KW Yeast; ss; gene; microbiological; prenyl alcohol; geranylgeraniol;
 KW farnesol; nerolidol; squalene synthetase inhibitor; BTS1.
 XX
 OS Saccharomycetes cerevisiae.
 XX
 Key Location/Qualifiers
 FT 1..1008
 CDS /*tag= a
 FT /product= "BTS1"
 XX
 PN JP2002300896-A.
 XX
 PD 15-OCT-2002.
 XX
 PF 18-JAN-2002; 2002JP-00010528.
 XX
 PR 30-JAN-2001; 2001JP-00021547.
 XX
 PA (TOYT) TOYOTA JIDOSHA KK.
 XX
 WI WI; 2003-485780/46.
 DR P-PSDB; ADC49388.
 DR
 PT A process for preparation of prenyl alcohols by culture of their
 PT productive microorganisms in a medium with added squalene synthetase
 PT inhibitor.
 XX
 PS Disclosure; SEQ ID NO 12; 37pp; Japanese.
 XX
 CC The invention discloses a microbiological preparation of prenyl alcohols,
 CC particularly geranylgeraniol, farnesol and/or nerolidol, by culture of
 CC their productive microorganism spp. of Saccharomycetes, Saccharomycopsis,
 CC Saccharomycodes, Schizosaccharomycetes, Wickerhamia, Debaryomyces,
 CC Hanseniaspora, Pichia, Candida, Zygosaccharomycetes, Ogataea, Kurashia,
 CC Komagataella, Yarrowia, Kluyveromyces, Trichosporon, Cryptococcus,
 CC Torulaspora, Bacillus, Staphylococcus, Micrococcus, Exiguobacterium,

CC	Query Match	Score	DB	Length
CC	Giberella or Mucor, or Ambrosiozyma, Cystofilobasidium, Metschnikowia, Trichosporiella, Xanthophyllomyces, Bullera, Fellomyces, Filobasidium, CC	64.2%;	1008;	1008;
CC	Holtermannia, Phaffia, Rhodotula, Sporidiobolus, Sporobolomyces, CC	100.08;	1008;	1008;
CC	Williopsis, Zygoascus, Leucosporidium, Myxozyma, Kloeckera, CC	0;	0;	0;
CC	Brevibacterium, Alcaligenes or Haloferax in a medium with added squalene CC	0;	0;	0;
CC	synthetase inhibitor to extra- and intra-cellular cumulate premyl CC	0;	0;	0;
CC	alcohols, followed by their collection. The sequence presented is the CC	0;	0;	0;
XX	yeast HMGI cDNA.	0;	0;	0;
SQ	Sequence 1008 BP; 337 A; 199 C; 176 G; 296 T; 0 U; 0 Other;	0;	0;	0;
QY	301 ATGAGGCGCAAGATAGATGAGCTCATCAATAATGATCTGTTTGGTCCAGCCAAATGAA 360			
DB	1 ATGAGGCCAAGATAGATGAGCTGATCAATGATCTGTTTGGTCCAGCCAAATGAA 60			
QY	361 AGCTTGATTTCAAAACCTTATAATCAATCATCTTTTGAAACCTGGCAAGACTTTAGACTA 420			
DB	61 AGCTTGATTTCAAAACCTTATAATCAATCATCTTTTGAAACCTGGCAAGACTTTAGACTA 120			
QY	421 AATTTAATAGTTCAAATTAACAGAGTTATGAAATTTGCCCAAGACCAGCTGGCCATAGTT 480			
DB	121 AATTTAATAGTTCAAATTAACAGAGTTATGAAATTTGCCCAAGACCAGCTGGCCATAGTT 180			
QY	481 TCGCAAAATTTGTTGAGCTCTTGCATAAATTCAGAGCCCTTTTAATCGCAGATATAGAGATAAT 540			
DB	181 TCGCAAAATTTGTTGAGCTCTTGCATAAATTCAGAGCCCTTTTAATCGCAGATATAGAGATAAT 240			
QY	541 GCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTAACCTCCACTATA 600			
DB	241 GCTCCCTTGAGAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTAACCTCCACTATA 300			
QY	601 AACACCGCAAAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCAAAA 660			
DB	301 AACACCGCAAAATATATGTAATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCAAAA 360			
QY	661 GAGCCTTTGTATCATAAATTTGATTACCAATTTTCAACGAAGAATTTGATCAATCTACATAGG 720			
DB	361 GAGCCTTTGTATCATAAATTTGATTACCAATTTTCAACGAAGAATTTGATCAATCTACATAGG 420			
QY	721 GGAAGAGCCTTGATATATATCTGAGAGACTTTCTGCTGAAATCAATCACTTACTCAGGAG 780			
DB	421 GGAAGAGCCTTGATATATATCTGAGAGACTTTCTGCTGAAATCAATCACTTACTCAGGAG 480			
QY	781 ATGTATTGTAATGTTTATCAATAAACAGGCGGCTTTTCAGATTAAAGTTGAGACTC 840			
DB	481 ATGTATTGTAATGTTTATCAATAAACAGGCGGCTTTTCAGATTAAAGTTGAGACTC 540			
QY	841 ATGGAAGCGCTGTCTCTCTCTCACACAGGCGCAATTCGTTGGTTTCCTTTTCAATAAATCTTT 900			
DB	541 ATGGAAGCGCTGTCTCTCTCTCACACAGGCGCAATTCGTTGGTTTCCTTTTCAATAAATCTTT 600			
QY	901 CTGGGTATTTATTCAGATTAGAGATGATTAATTGAAATTTGAAAGATTTCCAAATCTCC 960			
DB	601 CTGGGTATTTATTCAGATTAGAGATGATTAATTGAAATTTGAAAGATTTCCAAATCTCC 660			
QY	961 AGCGAAAAAGGCTTTGCTGAGGACATTCAGAGGGGAGTTATCTTTTCCCATCGTCCAC 1020			
DB	661 AGCGAAAAAGGCTTTGCTGAGGACATTCAGAGGGGAGTTATCTTTTCCCATCGTCCAC 720			
QY	1021 GCCTTTAACTTCACTAAACGAAAGGTCAAACTGAGCAACAAATGAAATTTCTAAGAATTT 1080			
DB	721 GCCTTTAACTTCACTAAACGAAAGGTCAAACTGAGCAACAAATGAAATTTCTAAGAATTT 780			
QY	1081 CTCCTGTTGAGGACAAGTGATAAAGATATAAACTAAAGCTGATTCAAATFACCTGGAATTC 1140			
DB	781 CTCCTGTTGAGGACAAGTGATAAAGATATAAACTAAAGCTGATTCAAATFACCTGGAATTC 840			
QY	1141 GACACCAATTCATTGGCTTACACCAAAATTTTATTAATCAATTAGTGAATATGATAAA 1200			

XX
SQ Sequence 599 BP; 193 A; 129 C; 106 G; 171 T; 0 U; 0 Other;
Query Match 36.2%; Score 568.6; DB 10; Length 599;
Best Local Similarity 98.5%; Pred. No. 1e-124;
Matches 574; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 299 CAATGGAGGCCAAGATAGATGAGCTGATCAATATGATCCTGTTGGTCCAGCCAAAATG 358
DB 17 CCAATGGAGGCCAAGATAGATGAGCTGATCAATATGATCCTGTTGGTCCAGCCAAAATG 76
QY 359 AAAGCTTGATTTTCAAAACCTTTAATACATCCATCTTTTGAACCTGGCAAGAACTTTAGAC 418
DB 77 AAAGCTTGATTTTCAAAACCTTTAATACATCCATCTTTTGAACCTGGCAAGAACTTTAGAC 136
QY 419 TAAATTTAATAGTTCAATTAACAGAGTTATGATTTGCCCCAAGACAGCTGGCCATAG 478
DB 137 TAAATTTAATAGTTCAATTAACAGAGTTATGATTTGCCCCAAGACAGCTGGCCATAG 196
QY 479 TTTGCGAAATGTTGAGCTCTTGATATTAATCCAGCCTTTTAAATCGAGATATAGAATA 538
DB 197 TTTGCGAAATGTTGAGCTCTTGATATTAATCCAGCCTTTTAAATCGAGATATAGAATA 256
QY 539 ATGCTCCCTTGAGAGGAGGACAGACCACTTCTCACTTAATCTTCGGTGTACCCCTCCACTA 598
DB 257 ATGCTCCCTTGAGAGGAGGACAGACCACTTCTCACTTAATCTTCGGTGTACCCCTCCACTA 316
QY 599 TAAACACCGCAATATATATGATTTTCAAGCCATGAGCCATGCACTGTATCGAGCTTAACCA 658
DB 317 TAAACACCGCAATATATATGATTTTCAAGCCATGAGCCATGCACTGTATCGAGCTTAACCA 376
QY 659 AAGAGCCTTTGATATATATGATTTTCAAGCCATGAGCCATGCACTGTATCGAGCTTAACCA 718
DB 377 AAGAGCCTTTGATATATATGATTTTCAAGCCATGAGCCATGCACTGTATCGAGCTTAACCA 436
QY 719 GGGGCAAGGCTTGGATATATATGATTTTCAAGCCATGAGCCATGCACTGTATCGAGCTTAACCA 778
DB 437 GGGGCAAGGCTTGGATATATATGATTTTCAAGCCATGAGCCATGCACTGTATCGAGCTTAACCA 496
QY 779 AGATGATTTTGAATATGATTTTCAAGCCATGAGCCATGCACTGTATCGAGCTTAACCA 838
DB 497 AGATGATTTTGAATATGATTTTCAAGCCATGAGCCATGCACTGTATCGAGCTTAACCA 556
QY 839 TCATGGAAGCGCTGTCT 881
DB 557 TCATGGAAGCGCTGTCT 599

RESULT 7
ID ADF75080
XX ADF75080 standard; DNA; 963 BP.
XX ADF75080;
AC
XX
DT 26-FEB-2004 (first entry)
XX
DE A_gossypii DNA essential for fungal growth & development SeqID 21.
XX
KW fungal growth; fungicide; screening assay; fungicidal; crop; gene; ds.
XX
OS Bremothecium gossypii.
XX
PN WO2003072706-A2.
XX
PD 04-SEP-2003.
XX
PF 11-FEB-2003; 2003WO-US004154.
XX
PR 22-FEB-2002; 2002US-0359557P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Gaffney TD, Flavien A, Dietrich F, Philippsen P, Kirksey CMW;

PI Gates K, Gardner J;
XX WPI; 2003-721756/68.
DR P-PSDB; ADF75081.
XX
PT Identifying a fungicidal compound comprising combining a polypeptide with
PT the compound to be tested with the ability to bind, or to inhibit the
PT activity of the polypeptide under conditions conducive to binding or
PT inhibition.
XX
PS Claim 11; SEQ ID NO 21; 127pp; English.
XX
CC This invention relates to novel nucleic acid molecules, isolated from
CC Aeshba gossypii, which encode proteins essential for fungal growth and
CC development. Specifically, it refers to methods for identifying these
CC proteins as fungicide targets by using gene disruption with a genetic
CC resistance gene. In particular, the nucleic acids encoding the proteins
CC required for fungal growth provide the basis of screening assays designed
CC to easily and rapidly identify inhibitory compounds that exhibit
CC fungicidal activity. Accordingly, the present invention describes such
CC methods to identify protein inhibitors that can be used as fungicides to
CC suppress the growth of undesirable fungi in agronomically important crops
CC such as maize, wheat, barley, cotton and sugar beet. This polynucleotide
CC sequence is a DNA sequence essential for fungal viability, used in an
CC exemplification of the invention.
XX
SQ Sequence 963 BP; 206 A; 252 C; 293 G; 212 T; 0 U; 0 Other;
Query Match 11.7%; Score 183.4; DB 10; Length 963;
Best Local Similarity 54.7%; Pred. No. 2.2e-33;
Matches 458; Conservative 0; Mismatches 361; Indels 18; Gaps 4;
QY 374 AACCTTATATACATCTCTTTGAAACCTGGCAAGAACTTTAGACTAAATTTAATAGTTC 433
DB 71 AGCTTTATATACATCTCTTTGAAACCTGGCAAGAACTTTTCCGCAATACGCTGATCCGCG 130
QY 434 AAATTAACAGATTAATGAATTTGCCAAAGACAGCTGGCCATAGTTTTCGCAATTTCTTG 493
DB 131 TGTTCAACGGGTTCTATGGTCTTTCCGAGCGGAGGTCGCGCGGTCAAGAGTTGTGTGG 190
QY 494 AGCTTTTGCAATAATCCAGCTTTTAAATCGACGATATAGAGATTAATGCTCCCTTGAGAA 553
DB 191 AGATGTTGCAATGCGGAGCTCTGTGATCGACGATATCGAAGACAACTCTGCTAGGAGAC 250
QY 554 GGGGACAGACCACTTCTCACTTAATCTTCTGTTGATACCTCCATTAACACCGCAATTT 613
DB 251 GTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 310
QY 614 ATATGATTTTCAGAGCATGCAACTTTGATCGCAGCTAAACCAAAA-----AGAGCCTT 667
DB 311 ACATGATTTTGTGTCATGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 370
QY 668 TGATCATATTTGATTAACGATTTTCAACGAAGAATTTGATCAATCTACATAGGGGCAAG 727
DB 371 CGCTGAGGATCTGCTGAAGTCTTCAACGAGGAGATGATGAATCTGCAATCGCGGGCAGG 430
QY 728 GCTTGTATATATCTGAGAGACTTTCTGCTGAAATATATACCTACTCTAGAGAGATTT 787
DB 431 GCCTTGACATATATTTGGCGGACACGCTTTACGG-----TGCGTGGGAGCATCACTATCC 484
QY 788 TGAATATGTTATGATAAAACAGCGGCTTTTTCAGATTAACGTTGAGACTCATGGAAG 847
DB 485 TCGGATGTTGATGATATAGACCGCGGCTTTGTTCCGGCTGACCGGTGCGGTATCATGGAAG 544
QY 848 CGCTGTCTCTCTCTCACA---CCAGCGGCATTTGTTGTTTCTTTTCATATAAATCTTCTGG 904
DB 545 CGTTACGGGAGGCCCCAGATGGACCGGATCAACATTTGTTGCGCTTAGCAACTGCTTG 604
QY 905 GTATATTTATCAGATTAAGATGATTTGATTTGAAAGATTTCCAAATTTCCAGCG 964
DB 605 GCGTGTGTACCAAGTGGCGGAGCACTATTGAACTCCGACCGCACTGCGCATCCGAAA 664
QY 965 AAAAGGCTTTGCTGAGGACATTAACAGAGGGAAGTTATCTTTTCCATCGTCCAGCCC 1024

```

Db 665 ACAAGGCTTTGACAGATATCATCCGAGGCAAAATTCCTACCCATTATCCAGGCC 724
Qy 1025 TTAATCTCACTAAACGAAAGTCAACTGAGCAACACAAATGAATTCATGAATCTCC 1084
Db 725 TGCAGTACGCCGCTGTCCA---TGACCCAGCGGCTACGACTTCCTGGTCACTGTGTAC 781
Qy 1085 TGTGAGACAGTGAATGAATATAAATAAAGCTGATCAATATCTGGAATTCGACA 1144
Db 782 GCGAGAGCAGACAGATATCACCAAGCGCGGTGGTGGATATCTTGGGACGTCA 841
Qy 1145 CCAATTCATTGGCCCTACCAAAAAATTTTATTAATCAATTAAGTGAATGATAAAA 1201
Db 842 GTGGTTCGCTCGGTACAGGAGCAGCTATCATCGAACTAGCAACAATTGATAAGA 898

RESULT 8
AAT31846
ID AAT31846 standard; cDNA; 903 BP.
XX
AC AAT31846;
XX
DT 17-FEB-1997 (first entry)
DE
DE Human geranylgeranyl pyrophosphate synthetase (GGPPS) cDNA sequence.
KW Geranylgeranyl pyrophosphate synthetase; cholesterol metabolism; steroid;
KW post-translational modification; gene therapy; rhabdomyolysis;
KW cell morphology; neoplasia; cancer; ss.
XX
OS Homo sapiens.
FH Key
CDS 1..903
/*tag= a

WO9621736-A1.
XX
XX 18-JUL-1996.
XX
XX 11-JAN-1995; 95WO-US000421.
XX
XX 11-JAN-1995; 95WO-US000421.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Greene JM, Rosen CA, Kirkness EP;
PI
PI WPI; 1996-342289/34.
DR P-PSDB; AAR97565.
XX
XX Polynucleotide encoding human geranyl:geranyl pyrophosphate synthetase -
PT fragments of which can be used as probes to diagnose disease or
PT susceptibility to disease related to mutation in above polynucleotide.
XX
XX Claim 7; Page 38; 55pp; English.
XX
XX AAT31846 is a cDNA sequence encoding human geranylgeranyl pyrophosphate
XX synthetase (GGPPS). The sequence was obtained from a foetal heart cDNA
XX library and identified by homology to the corresponding enzyme of
XX Neurospora crassa. The GGPPS enzyme is involved in post-translational
XX modification (more specifically the geranylgeranylation) of proteins and
XX is also involved in a branch of the cholesterol/steroid metabolic
XX pathway. GGPPS catalyzes the trans-addition of 3 molecules of isopentenyl
XX diphosphate onto dimethylallyl diphosphate to form C20 GGPP. GGPP
XX controls the signal transduction, activation and trafficking of GGPP-
XX modified proteins. Human GGPPS can be used to control cell morphology
XX e.g. to treat rhabdomyolysis (a side effect of treatment with HMG-CoA
XX reductase) and unwanted apoptosis. Antagonists of GGPPS can be used to
XX treat neoplasms and to prevent viral proliferation, increased levels of
XX GGPPS are indicative of cancer cell growth
XX
XX Sequence 903 BP; 305 A; 173 C; 183 G; 242 T; 0 U; 0 Other;
SQ

```

```

Query Match
Best Local Similarity 7.3%; Score 114; DB 2; Length 903;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;
Qy 359 AAAGCTTGATTTCAAAACCTTATATACATCCCTTTTGAACCTGGCAAGACTTTAGAC 418
Db 26 AAAGAAATCTTCTAGAACCTTATAAATACTTACTTCACTTACCAAGTTAAACAAGTGAGAA 85
Qy 419 TAAATTTAAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATAG 478
Db 86 CCAAACTTTTCAAGGCATTTAATCATTTGGCTGAAAGTTCCAGAGGACAAAGCTACAGATA 145
Qy 479 TTTCGCAAAATGTTTCAGCTCTTTGCATATAATTCAGAGCTTTTAAATCGACGATATAGAATA 538
Db 146 TTATTTGAAGTGACAGAAATGTTGCATATGCCAGTTTACTCATCGATGATATGAAGACA 205
Qy 539 ATGCTCCCTTGAGAAAGGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTCCACTA 598
Db 206 ACTCAAAACTCCGACGTGGCTTTCCAGTGGCCACACAGCATCTATGGAATCCCATCTGTCA 265
Qy 599 TAAACACCGCAAAATTAATATGATATTTTCAGAGCCATGCAACTTGTATCGCAGCTTAACCAA 658
Db 266 TCAATCTGCCAAATTAACGTGATTTTCCTTGGCTTGGAGAAAGTCT----- 310
Qy 659 AAGAGCCTTTGTATCATATAATTTGATTCAGATTTTCAACGAAAGATTTGATCAATCTACATA 718
Db 311 TAAACCTTTGATCACCAGATGCAGTGAAGCTTTTACCGCCAGCTTTTGGAACTCCATC 370
Qy 719 GGGGCAAGGCTTGATATATATCTGAGAGACTTCTGCTGGAATCATACCTACTCAGG 778
Db 371 AGGCAAGGCTCTAGATATTTTACTGGAGGGA-----TAAATTACACTTTGTCCCACTGAAG 424
Qy 779 AGATGTTTGAATATGTTATGAATAAAACAGGGCGCTTTTTCAGATTAAAGTTGAGAC 838
Db 425 AAGAATATAAGCTTAAGTGTGTCGAGAAAACAGGTGGACTGTTTGGATTAGCAGTAGGTC 484
Qy 839 TCATGGAAGCGCTGTCTCCTTCTCACACCGGCCATTCGTTGGTTCCTTTTCATAAAATC 898
Db 485 TCATGCACTGTCTCTGATTACAAAGAAAGATTTAAACCGCTACTT-----AATA 535
Qy 899 TTCTGGGTATTTATTCAGATTAGAGATGATTTCAATTTGAAAGATTTCCAAATGT 958
Db 536 CACTTGGGCTCTTTTTCAAAATTAGGATGATTTATGCTAAATCATACATCCCAAGAAATATA 595
Qy 959 CCAGCGAAAAGGCTTTCGTGAGGACATTACAGAGGGAAGTTATCTTTTCCCATCTGCC 1018
Db 596 GTGAAAACAAAAGTTTGGTGAAGATCTGACAGAGGGAAGTTCTCATTTCTCTACTATTC 655
Qy 1019 ACGCCCTT 1026
Db 656 ATGCTATT 663

RESULT 9
AAV41555
ID AAV41555 standard; cDNA; 903 BP.
XX
AC AAV41555;
XX
DT 12-OCT-1998 (first entry)
DE Human geranyl geranyl pyrophosphate synthase.
XX
XX Human; geranyl pyrophosphate synthase; hGGPS; antagonist;
XX agonist; ss.
XX
XX Homo sapiens.
FH Key
CDS 1..903
/*tag= a
/product= "human geranyl geranyl pyrophosphate synthase"
FT
FT

```

XX US5786193-A.
 XX PD 28-JUL-1998.
 XX PF 06-JUN-1995; 95US-00469665.
 XX PR 06-JUN-1995; 95US-00469665.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Rosen CA, Greene JM, Kirkness EF;
 XX DR WPI; 1998-436529/37.
 XX DR P-PSDB; AAW59753.
 XX PT DNA encoding human geranyl pyrophosphate synthetase polypeptide -
 XX PT useful for producing recombinant polypeptide.
 XX PS Claim 5; Fig 1; 22pp; English.
 XX CC This is the nucleotide sequence of the human geranyl geranyl
 CC pyrophosphate synthase (hGPPS) used in the method of the invention. The
 CC proteins are used in the production of recombinant hGPPS, which can be
 CC used to screen for hGPPS antagonists or agonists
 XX SQ Sequence 903 BP; 304 A; 173 C; 184 G; 242 T; 0 U; 0 Other;
 Query Match 7.3%; Score 114; DB 2; Length 903;
 Best Local Similarity 52.1%; Pred. No. 6.1e-17;
 Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;
 QY 359 AAAGCTTGGATTCAAAACCTTATATACATCCTTTTGAACCTGGCAAGACTTTAGAC 418
 DB 26 AAGAATCTCTTAGAACCTTATATATCTTCTAGTTACCAAGTAAACAAGTGAGAA 85
 QY 419 TAAATTTAATAGTTCAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGCCATAG 478
 DB 86 CCAAACTTTTCCAGGCAATTTAATCATTTGGCTGAAAGTTCCAGAGGACAAGCTACAGATTA 145
 QY 479 TTTCCGAAATTTGTAGCTCTTCATTAATTTCCAGCCCTTTTAAATCGAGATATAGAGATA 538
 DB 146 TTATTGAAGTGACAGAAATGTTGCATAATGCCAGTTTACTCATCGATGATATTGAAGACA 205
 QY 539 ATGCTCCCTTGGAGAAGGGGACAGACCTTCTCCTTAACTTAACTTCGGTGTACCTCCACTA 598
 DB 206 ACTCAAACTCCGACGGCTTTCCAGTGGCCCAAGCACTATGGAATCCCATCTGTCA 265
 QY 599 TAAACACCGCAATATATATGTTATTCAGAGCCATGCACTTTGTATCGCAGCTAACCAAA 658
 DB 266 TCAATTCGCAATTTACGTGTATTTCTTGGCTTGGAGAAAGTCT----- 310
 QY 659 AAGAGCTTTGTATCATATTTGATTTAGATTTTCAACGAGAAATGATCAATCTACATA 718
 DB 311 TAACCTTTGATACCCAGATGCACTGAAGCTTTTACCCGCGAGCTTTTGGAACTCCATC 370
 QY 719 GGGGACAAGGCTTGGATATATCTAGGAGAGACTTTCTGCTGTAATCATACCTACTCAGG 778
 DB 371 AGGACAAGGCTTGGATATTTACTTGGAGGGA-----TAATTAACCTTGTCCCACTGAAG 424
 QY 779 AGATGTTATTTGAATGTTATGTAATAAAACAGCGGCGCTTTTTCAGATTAACAGTTGAGAC 838
 DB 425 AAGAATATAAGCTATGTTGCTGTCAGAAACAGGTGGACTGTTTGGATTAGCAGTAGTC 484
 QY 839 TCATGGAAGCGCTGTCTCTTCCCTCAACCAAGCGCCATTTGTTGGTTCCCTTTTCAATAATC 898
 DB 485 TCATGCAAGTTGTTCTCTGATTAACAAAGAGATTTAAACCGCTACTT-----AATA 535
 QY 899 TTCTGGGTATTTATTCAGATGAGATGATTAATTTGAATTTGAAGATTTCCAAATGT 958
 DB 536 CACTTGGGCTCTTTTCCAAATTAGGATGATTAATGCTAATCTACACTCCAAAGATATA 595
 QY 959 CCAGCGAAAAGGCTTTGCTGAGGACATTTACAGAGGGGAAAGTTATCTTTTCCCATGCTCC 1018

DB 596 GTGAAAACAAAGTTTGGGTGAAGATCTGACAGAGGGAAGTTCTCATTTCTACTATTC 655
 QY 1019 AGCCCTT 1026
 DB 656 ATGCTATT 663
 RESULT 10
 AAX86773
 ID AAX86773 standard; cDNA; 903 BP.
 AC AAX86773;
 XX 20-SEP-1999 (first entry)
 DT Human geranylgeranyl pyrophosphate synthetase (hGPPS) encoding cDNA.
 DE Geranylgeranyl pyrophosphate synthetase; hGPPS; cell morphology; human;
 KW apoptosis; neoplasia; tumour; cancer; viral proliferation; gene therapy;
 KW human Choroideremia; ss.
 XX Homo sapiens.
 OS US5928924-A.
 XX 27-JUL-1999.
 PD 11-MAR-1998; 98US-00038596.
 PF 11-JAN-1995; 95WO-US000421.
 PR 06-JUN-1995; 95US-00469665.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Kirkness EF, Greene JM;
 PI WPI; 1999-429504/36.
 DR P-PSDB; AAY27005.
 XX Geranylgeranyl pyrophosphate synthetase peptides useful for regulating
 PT apoptosis.
 PS Claim 18; Fig 1A-D; 23pp; English.
 CC This cDNA encodes a human geranylgeranyl pyrophosphate synthetase (hGPPS)
 CC polypeptide. The hGPPS polynucleotides and polypeptides may be used to
 CC control the morphology of cells, suppress apoptosis, screen for hGPPS
 CC agonists and antagonists (which may in turn be used to treat neoplasia,
 CC e.g. tumours and cancer cell growth, and prevent viral proliferation),
 CC gene therapy, and diagnose a disease or susceptibility to diseases
 CC related to a mutation in hGPPS (e.g. human choroideremia)
 XX SQ Sequence 903 BP; 304 A; 173 C; 184 G; 242 T; 0 U; 0 Other;
 Query Match 7.3%; Score 114; DB 2; Length 903;
 Best Local Similarity 52.1%; Pred. No. 6.1e-17;
 Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;
 QY 359 AAAGCTTGGATTCAAAACCTTATATCAGATCTTGTGAACCTGGCAAGACTTTAGAC 418
 DB 26 AAGAATCTCTTAGAACCTTATATATCTTCTAGTTACCAAGTAAACAAGTGAGAA 85
 QY 419 TAAATTTAATAGTTCAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGCCATAG 478
 DB 86 CCAAACTTTTCCAGGCAATTTAATCATTTGGCTGAAAGTTCCAGAGGACAAGCTACAGATTA 145
 QY 479 TTTCCGAAATTTGTAGCTCTTCATTAATTTCCAGCCCTTTTAAATCGAGATATAGAGATA 538
 DB 146 TTATTGAAGTGACAGAAATGTTGCATAATGCCAGTTTACTCATCGATGATATTGAAGACA 205
 QY 539 ATGCTCCCTTGGAGAAGGGGACAGACCTTCTCCTTAACTTAACTTCGGTGTACCTCCACTA 598

Db 206 ACTCAAACTCGAGCTGGCTTCAGTGGCCACAGCATCTATGGAATCCCATCTGTCA 265
QY 599 TAACACCGCAAAATATATATATATTTTCAGAGCATCAACTGTATCGCAGCTTAACACA 658
Db 266 TCAATTTCTGCCAAATACGTGTATTTCTTGGCTTGGAGAAAGTCT- 310
QY 659 AAGAGCCTTTGTATCATATATTTGATTACGATTTTCAACGAAGATTTGATCAATCTACATA 718
Db 311 TAACCCCTTGATCACCAGATGAGTGAAGCTTTTACCGCAGCTTTTGGAACTCCATC 370
QY 719 GGGGACAGGCTTGTATATATATCTACTGAGAGACTTTCTGCCTGAAATCATACCTACTCAGG 778
Db 371 AGGACAAAGGCTAGATATTTACTGGAGGA-----TAATTACACTTGTCCCACTGAAG 424
QY 779 AGATGTATTTGAATATGTATTGATTAATAAACAAGGCGCTTTTCAGATTAACGTTGAGAC 838
Db 425 AAGAATATAAAGCTATGGTCTGCGAGAAACAGGTGGACTGTTGGATTAGCAGTAGGTTC 484
QY 839 TCATGGAAGCGCTGTCTCCTCTCACACACCGGCATTCGTTGGTTCCTTTTCATAAATC 898
Db 485 TCATGCAAGTGTCTCTGATTACAGAGAGATTTAAACCGCTACTT-----ATA 535
QY 899 TTCTGGGTATTTATCATAGATTAGAGATGATTACTTGAATTTGAAAGATTTTCCAAATGT 958
Db 536 CACTTGGGCTCTTTTCCAAATTAGGATGATTATGCTAATCTACACTCCAAAGATATA 595
QY 959 CCAGGAAAAGGCTTGTCTGAGGACATTTACAGAGGGAAGTATCTTTTCCCATCGTCC 1018
Db 596 GTGAAAACAAAAGTTTGGGTGAAGATCTGACAGAGGGAAGTTCTCATTTCTCTACTATTC 655
QY 1019 AGCCCTTT 1026
Db 656 ATGCTATT 663
RESULT 11
ADJ10255
ID ADJ10255 standard; DNA; 1395 BP.
XX AC ADJ10255;
XX DT 17-JUN-2004 (first entry)
XX DE Human geranylgeranyl diphosphate synthase 1 DNA SeqID 4.
XX KW gene; ds; chromosome 1q43; human; geranylgeranyl diphosphate synthase 1;
KW GGPS1; geranylgeranyl pyrophosphate synthetase; GGPPS; ggppsase;
KW geranyltransferase; embryonic development; cell differentiation;
KW apoptosis; developmental; hyperproliferative disorder; cancer;
KW cytosolic.
XX OS Homo sapiens.
XX FH Key
XX CDS 170..1072
XX FT /*tag= a
XX FT /product= "GGPS1 protein"
XX PN US2004005570-A1.
XX PD 08-JAN-2004.
XX PF 02-JUL-2002; 2002US-00189268.
XX PR 02-JUL-2002; 2002US-00189268.
XX (ISIS-) ISIS PHARM INC.
XX Dean NM, Bennett CF, Dobie KW;
XX WPI; 2004-201146/19.
XX P-PSDB; ADJ10383.
XX

PT New antisense oligonucleotides for modulating geranylgeranyl diphosphate
PT synthase 1 expression, useful for diagnosing, preventing or treating
PT conditions associated with the protein, e.g. cancer.
XX
XX Example 13; SEQ ID NO 4; 76pp; English.
XX This invention relates to a novel antisense compounds that modulate the
CC expression of human geranylgeranyl diphosphate synthase 1 (also known as
CC GGPS1, geranylgeranyl pyrophosphate synthetase, GGPPS, ggppsase and
CC geranyltransferase) and located on chromosome 1p43. Specifically, it
CC refers to compositions useful for inhibiting the expression of GGPS1,
CC which normally participates in embryonic development, cell
CC differentiation and stimulates apoptosis via caspase-3 activation. The
CC present invention describes antisense oligonucleotides that comprise at
CC least one modified sugar moiety, a 2'-O-methoxyethyl (2' MOE) and at
CC least one modified nucleobase, a 5-methylcytosine. Accordingly, these
CC compounds are useful for treating a disease or condition associated with
CC GGPS1 such as a developmental or hyperproliferative disorder (e.g.
CC cancer) that arise as a result of aberrant apoptosis. As such, these
CC compositions exhibit cytostatic activity and are useful for diagnostics,
CC prophylaxis, research reagents and various kits. This polynucleotide
CC sequence is the geranylgeranyl diphosphate synthase 1 DNA of the
CC invention.
XX
SQ Sequence 1395 BP; 449 A; 259 C; 286 G; 401 T; 0 U; 0 Other;
Query Match 7.3%; Score 114; DB 12; Length 1395;
Best Local Similarity 52.1%; Pred. No. 7e-17;
Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;
QY 359 AAAGCTTGATTTCAAACCTTATAATCATCATCTTTTGAACCTCGCAAGACTTTAGAC 418
Db 195 AAAGAATTTCTTAGAACCTTATAATAATCTTACTTACGTTACAGTAAACAAGTGAGAA 254
QY 419 TAAATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCAAAGACGCTGGCCATAG 478
Db 255 CCAAACTTTTCACAGGCAATTTAATCATTTGGCTGAAAGTTCCAGAGGACAGCTACGATTA 314
QY 479 TTTTCGCAATTTGTTGAGCTCTTGCATATTTCCAGCTTTTAATCCAGCATATAGAGATA 538
Db 315 TTATTGAAGTGACAGAAATGTTGCATATATGCCAGTTTACTCATCATGATATTTGAAGACA 374
QY 539 ATGCTCCTCTTGAGAAAGGGGACAGACCCTTCTCATTAAATCTTCGGTGTACCTCCACTA 598
Db 375 ACTCAAACTTCGACGCTGGCTTTCCAGTGGCCACAGCATCTATGAAATCCCATCTGTCA 434
QY 599 TAAACACCGCAAAATATATATATTTTCAGAGCATCACTTGTATCGAGCTTAACACAA 658
Db 435 TCAATTTCTGCCAAATTAACGTGATTTTCTTGGCTTGGAGAAAGTCT----- 479
QY 659 AAGAGCCTTTGTATCATATATTTGATTACGATTTTCAACGAAGATTTGATCAATCTACATA 718
Db 480 TAAACCTTGATCACCAGATGAGTGAAGCTTTTACCGCCAGCTTTTGGAACTCCATC 539
QY 719 GGGGACAGGCTTGGATATATATCTCGAGAGACTTTCTGTGCTGAAATCATACCTACTCAGG 778
Db 540 AGGACAAAGGCTAGATATTTACTGGAGGA-----TAATTACACTTGTCCCACTGAAG 593
QY 779 AGATGTATTTGAATATGTATTGATTAATAAACAAGGCGCTTTTTCAGATTAACGTTGAGAC 838
Db 594 AAGAATATAAAGCTATGGTCTGCGAGAAACAGGTGGACTGTTGGATTAGCAGTAGGTTC 653
QY 839 TCATGGAAGCGCTGTCTCCTCTCTCACACACCGGCATTCGTTGGTTCCTTTTCATAAATC 898
Db 654 TCATGCAAGTGTCTCTGATTACAAAGAGATTTAAACCGCTACTT-----ATA 704
QY 899 TTCTGGGTATTTATCATAGATTAGAGATGATTTCTTGAATTTGAAGATTTCCAAATGT 958
Db 705 CACTTGGGCTCTTTTCCAAATTAGGATGATTATGCTAATCTACACTCCAAAGATATA 764
QY 959 CCAGCGAAAAGGCTTTGCTGAGGACATTTACAGAGGGAAGTATCTTTTCCCATCGTCC 1018
Db 765 GTGAAAACAAAAGTTTTTGTGAAGATCTGACAGAGGGAAGTTCTCATTTCTCTACTATTC 824

XX 23-JUL-1999; 99WO-1B001353.
 XX 23-JUL-1998; 98US-0093940P.
 XX (GIST) GENSET.
 XX Bouguerel L;
 XX WPI; 2000-182704/16.
 XX P-PSDB; AAY68909.
 XX New isolated human geranyl-pyrophosphate synthetase nucleic
 XX acids, used to develop agents for the diagnosis of, e.g. pathologies
 XX related to a defect in the mevalonic biosynthetic pathway.
 XX Claim 3; Page 82-84; 88pp; English.
 XX The present sequence encodes a human geranylgeranyl pyrophosphate
 XX synthetase (hGGPPS). Two differently spliced mRNAs exist for this gene.
 XX The first spliced mRNA is derived from a cDNA (AAZ60888) which comprises
 XX exons 1, 2, 3 and 4. The second mRNA is derived from a cDNA (AAZ60889)
 XX which comprises 1b1s, 2, 3, and 4. The hGGPPS gene is located on
 XX chromosome 1, at the 1q42-1q43 locus. This chromosome 1 locus has been
 XX shown to carry a predisposing gene for prostate cancer. The nucleic acids
 XX encoding hGGPPS can be used for screening for agents which modulate the
 XX expression of the hGGPPS gene. Such agents can be used in therapeutic
 XX applications. The allelic markers associated with the hGGPPS gene can
 XX be used for the diagnosis of diseases related to an alteration in the
 XX regulatory or coding regions of hGGPPS, such as pathologies related to a
 XX defect in the mevalonic biosynthetic pathway. The products can also be
 XX used for detection, diagnosis and drug screening
 XX SQ Sequence 1547 BP; 484 A; 296 C; 315 G; 447 T; 0 U; 5 Other;
 Query Match 7.3%; Score 114; DB 3; Length 1547;
 Best Local Similarity 52.1%; Pred. No. 7.2e-17;
 Matches 348; Conservative 0; Mismatches 290; Indels 30; Gaps 3;
 QY 359 AAGCTTGGATTCAAACTTATATACATCTTTTGAACCTGGCAGACTTTAGAC 418
 DB 243 AAGAATTTCTTGAACCTTATATATCTTACCTGTTACCGATTAACCAAGTGAGAA 302
 QY 419 TAAATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCAAAGACAGCTGGCCATAG 478
 DB 303 CCAACTTTTCAGGCAATTAATCATCTGGCTGAAGTTCCAGAGCAAGCTACAGATTA 362
 QY 479 TTTCGCAATTTGAGCTTTGCAATATTCAGGCTTTTATCGACGATATAGAGATA 538
 DB 363 TTATTGAAGTGACAGAAATGTTGCATAATGCCAGTTTACTCATCATGATGATTTGAAGACA 422
 QY 539 ATGCTCCCTTGAGAGGGGACAGACCACTTCTCATTATCTCGTGTACCTCCACTA 598
 DB 423 ACTAAACTCGAGCTGGCTTTCCAGTGCGCCACAGCATCTATGAATCCCATCTGTCA 482
 QY 599 TAAACACGCGCAATATATATATTTTACAGAGCCATGCACTTGTATCGAGTAAACACAA 658
 DB 483 TCAATCTGCGCAATACGTTGATTTCTTGGCTTGGAGAACTCT----- 527
 QY 659 AAGAGCTTTGATCATATATTTGATGATGATTTTCAAGAGAAATGATCAATCATACATA 718
 DB 528 TAAACCTTTGATCACCAGATGAGTGAAGCTTTTACCAGGCTTTTGAAGCTCCATC 587
 QY 719 GGGGACAGAGCTTGGATATATCTCGAGAGACTTTCTGCTCAATCATACCTACTCAGG 778
 DB 588 AGGACAGAGCTTGAATATTTACTGAGGGA-----TAATACACTGTCCCATCTGAG 641
 QY 779 AGATGATTTGATATGTTATGATATGATATGATATGATATGATATGATATGATATGAT 838
 DB 642 AAGAATATGAAGCTATGTTGCTGCGAGAAACAGGTGGAGCTTTTGGATAGCAGTAGGTC 701
 QY 839 TCATGGAAGCGTGTCT 898

DB 702 TCATGCAAGTTGTTCTCTGATTACAAAGAAATTTAAACCGCTACTT-----AATA 752
 QY 899 TTCGGGCTATTTATTTATCAGATTACAGATGATATCTTGAATTTGAAGATTTCCAAATCT 958
 DB 753 CACTTGGGCTCTTTTCCAAATTAGGATGATATGCTAATCTACACTCCAAAGATATA 812
 QY 959 CCAGCGAAAGAGGCTTTGCTGAGGACATTACAGAGGGAAGTTATCTTTTCCCATCTGTC 1018
 DB 813 GTGAAAACAAAGATTTTGTGAGATCTCAGAGAGGGAAGTTCTCATTTCTCTACTATTC 872
 QY 1019 ACGCCCTT 1026
 DB 873 ATGCTATT 880
 RESULT 14
 ID ADL63550 standard; DNA; 1692 BP.
 AC ADL63550;
 XX 20-MAY-2004 (first entry)
 DE Human ovarian cancer DNA marker #21762.
 KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
 OS Homo sapiens.
 XX WO200170979-A2.
 PD 27-SEP-2001.
 XX 21-MAR-2001; 2001WO-US009126.
 XX 21-MAR-2000; 2000US-0191031P.
 PR 25-MAY-2000; 2000US-0207124P.
 PR 15-JUN-2000; 2000US-0211940P.
 PR 07-JUL-2000; 2000US-0216820P.
 PR 25-JUL-2000; 2000US-0220661P.
 PR 21-DEC-2000; 2000US-0257672P.
 (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 Lee J, Lillie J;
 WPI; 2001-611502/70.
 Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
 cancer cells as compared to their normal non-cancerous ovarian cells are
 used to characterize stage, grade, histological type of ovarian cancer.
 Disclosure; SEQ ID NO 21762; 106pp; English.
 The invention relates to nucleic acid markers which are overexpressed in
 ovarian cancer cells as compared to their expression in normal (i.e. non-
 cancerous) ovarian cells. The invention also relates to polypeptides
 encoded by the markers, antibodies that selectively bind to the
 polypeptides, a method of inhibiting ovarian cancer in a patient at risk
 of developing ovarian cancer involving inhibiting expression of a gene
 corresponding to a marker of the invention and a method of treating a
 patient afflicted with ovarian cancer comprising providing to cells of
 the patient an antisense oligonucleotide complementary to a marker of the
 invention. The markers are useful for assessing if a patient is afflicted
 with ovarian cancer, which involves comparing the level of expression of
 a marker in a patient sample and a normal level of expression of the
 marker in a control non-ovarian cancer sample. A difference between the
 expression levels indicates ovarian cancer. The level of expression of a
 marker corresponds to a secreted protein or to a transcribed
 polynucleotide or its portion. The level of expression of the marker is
 assessed by detecting the presence in the sample, a protein or protein
 fragment corresponding to the marker. The presence of protein or protein
 fragment is detected using an antibody that specifically binds with the

Db 378 TTATTGAAGTGACAGAAATGTTGCATAAATGCCAGTTTACTCATCGATGATATTGAAGACA 437
 QY 539 ATGCTCCCTTGTAGAGGGGACAGACCACCTTCTCAGCTTATCTTCGGTGTACCTCCACTA 598
 Db 438 ACTGAAACTCCGACGTGGCTTTCAGTGGCCACAGCACTATGGAATCCCACATCTGTCA 497
 QY 599 TAAACACCGCAAAATATATGTATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCCACAA 658
 Db 498 TCAATTCGCCAATTAACGTGTATTTCTTGGCTTGGAGAAAGTCT-----542
 QY 659 AAGAGCCTTTGTATCAATAATTGATTAACGATTTTCAACGAGAAATGATCAATCTACATA 718
 Db 543 TAAACCTTTGATCACCCAGATGAGTGAAGCTTTTACCCGCCAGCTTTTGGAACTCCATC 602
 QY 719 GGGGACAGGCTTTGGATATATATCTGAGAGACTTTCTGCTGAAATCATACCTACTCAGG 778
 Db 603 AGGGACAGGCTTAGAATTTTACTGGAGGA-----TAAATACACTTGTCCCACCTGAAG 656
 QY 779 AGATGTATTTGAATATGTTTATGAATAAAACAGGGGCGCTTTTTCAGATTAAACGTTGAGAC 838
 Db 657 AAGAAATAAAGCTATGTTGCTGCGAGAAACAGGTGACTGTTTGGATTAGCAGTAGGTC 716
 QY 839 TCATGGAAGCGTGTCTCTCTCTCACACACGGCGATTCGTTGGTTCCTTTTCATAAATC 898
 Db 717 TCATGCAAGTGTCTCTGATTACAAAGAAAGATTTAAACCGCTACTT-----AATA 767
 QY 899 TTCTGGGTATTATTATATCAGATTAGAGATGATTACTTGAATTTGAAAGATTTCCAAATGT 958
 Db 768 CACTTGGGCTCTTTTCCAAATTAGGATGATTATGCTTAATCTACATCCAAAGATATA 827
 QY 959 CCAGCGAAAGAGCTTTGCTGAGACATTACAGAGGGGAAGTTATCTTTTCCCATCGTCC 1018
 Db 828 GTGAAAACAAAGTTTTCGTAAGATCTGACAGAGGGAAGTTCTCAITTTCTACTATTC 887
 QY 1019 ACGCCCTT 1026
 Db 888 ATGCTATT 895

Search completed: October 16, 2004, 23:38:17
 Job time : 738.231 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 18, 2004, 07:18:36 ; Search time 0.001 Seconds
(without alignments)
1051.230 Million cell updates/sec

Title: us-10-041-018-22

Perfect score: 1720

Sequence: 1 MEAKIDELINDPVWSSQNE.....DTATNLHDELLYIIDHLSL 335

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1 seqs, 1569 residues

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=soft -Ous-10-041-018-22 -DB=us-10-041-018-1
-SUFFIX=pto -OUT=align22_1 -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bits
-START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=1 -DOCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=1 -MODE=LOCAL -OUTFMT=ptc
-NRM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -NCPU=6 -NO_XLPXY
-NEG_SCORES=0 -LONGLOG -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : us-10-041-018-1.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1720	100.0	1569	1	us-10-041-018-1

ALIGNMENTS

RESULT 1

us-10-041-018-1

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
0	1720.00	1569	335
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

us-10-041-018-22 (1-335) x us-10-041-018-1 (1-1569)

Qy	1	MetGluAlaLysIleAspGluLeuIleAsnAspProValTrpSerSerGlnAsnGlu	20
Db	301	ATGAGGCGCAAGATAGATGAGCTGATCAATATGATCCTGTTGGTCCGAGCAAAATGAA	360

Qy	21	SerLeuIleSerLysProTyrAsnHisIleLeuLeuLysProGlyLysAsnPheArgLeu	40
Db	361	AGCTTGATTTCAAACCTTATAATCATCTCTTTGAAACCTGGCAAGAACTTTAGACTA	420
Qy	41	AsnLeuIleValGlnIleAsnArgValMetAsnLeuProLysAspGlnLeuAlaIleVal	60
Db	421	AATTTAATAGTTCAAATTAACAGATTATGAATTTGCCAAAGACCAGCTGGCATAGTT	480
Qy	61	SerGlnIleValGluLeuLeuHisAsnSerSerLeuLeuIleAspAspIleGluAspAsn	80
Db	481	TGCAAAATGTTGAGCTCTTGCAATAATTCAGCCTTTTAAATCGACGATATAGAGATAAT	540
Qy	81	AlaProLeuArgArgGlyGlnThrThrSerHisLeuIlePheGlyValProSerThrIle	100
Db	541	GCTCCCTTGAGAAGGGGACAGACCACTTCTCACTTAATCTTGGTGTACCTCCACTATA	600
Qy	101	AsnThrAlaAsnTyrMetTyrPheArgAlaMetGlnLeuValSerGlnIleThrLys	120
Db	601	ACACCGCAAAATTAATGATTTATGAGCCATGCAACTTGTATCGCAGCTAACCAAAA	660
Qy	121	GluProLeuTyrHisAsnLeuIleThrIlePheAsnGluGluLeuIleAsnLeuHisArg	140
Db	661	GAGCCTTTGTATCATATAATTTGATTAGATTTTCAACGAGAATTTGATCAATCTACATAGG	720
Qy	141	GlyGlnGlyIleuAspIleTyrTrpArgAspPheLeuProGluIleIleProThrGlnGlu	160
Db	721	GGACCAAGGCTTGGATATATATCTGAGAGACTTCTCGCTGAAATCATACTACTCAGGAG	780
Qy	161	MetTyrLeuAsnMetValMetAsnLysThrGlyGlyLeuPheArgLeuThrLeuArgLeu	180
Db	781	ALGTATTTGATGATGTTATGATTAATAAAGCAGCGGCTTTTTCAGATTAACTGAGACTC	840
Qy	181	MetGluAlaLeuSerProSerSerHisGlyHisSerLeuValProPheIleAsnLeu	200
Db	841	ATGGAAGCGCTGCTCTCTTCTCACACACGCGCAATTCGTGGTTCCTTTCAATAATCTT	900
Qy	201	LeuGlyIleIleTyrGlnIleArgAspTyrLeuAsnLeuLysAspPheGlnMetSer	220
Db	901	CTGGGTATTTATATATCAGATTAGATGATTAATTTGAAATTTGAAAGATTTTCAAAATGTC	960
Qy	221	SerGluLysGlyPheAlaGluAspIleThrGluGlyLysLeuSerPheProIleValHis	240
Db	961	ACGAAAAAGGCTTTGCTGAGACATTACAGAGGGAAGTTATCTTTTCCCATCTGCCAC	1020
Qy	241	AlaLeuAsnPheThrLysThrLysGlyGlnThrGlnHisAsnGluIleLeuArgIle	260
Db	1021	GCCCTTAACCTTCACTAAACGAAAGGTCAAACTGAGCAACACATGAATTTCAAGATTT	1080
Qy	261	LeuLeuLeuArgThrSerAspLysAspIleLysLeuLysLeuIleGlnIleLeuGluPhe	280
Db	1081	CTCCTGTTGAGGACAAGTGATAAAGATATAAACTAAAGCTGATTCAATAATCTGGAATTC	1140
Qy	281	AspThrAsnSerLeuAlaTyrThrLysAsnPheIleAsnGlnLeuValAsnMetIleLys	300
Db	1141	GACACCAATTCATTGGCCCTACACCAAAATTTTATTAATCAATTAGTGAATATGATAAAA	1200
Qy	301	AsnAspAsnGluAsnLysTyrLeuProAspLeuAlaSerHisSerAspThrAlaThrAsn	320
Db	1201	AATGATAATGAAATAAGTATTTTACCTGATTTGGCTTCGCAATTCGACACCGCCCAAT	1260
Qy	321	LeuHisAspGluLeuLeuTyrIleIleAspHisLeuSerGluLeu	335
Db	1261	TTACATGACGAATTTGTTATATATATATAGACCCACTTATCCGAATTG	1305

Search completed: October 18, 2004, 07:18:37

Job time : 1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 18, 2004, 07:20:37 ; Search time 2 Seconds

(without alignments)

2.189 Million cell updates/sec

Title: us-10-041-018-383

Perfect score: 4113

Sequence: 1 MNLSLCIASPLLTksNPAA.....TVKDIIYNPLVLVNENEER 784

Scoring table: BLOSUM62

102

Searched: 1 seqs, 2792 residues

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1 summaries

Command line parameters:

```

-MODEL=frame p2n.model -DEV=soft -Q=us-10-041-018-383 -DB=us-10-041-018-361
-SUFFIX=pfo -OUT=align383_361 -MINWAT=0.1 -MOWCPU=0 -LOOPEXT=0 -UNITS=bits
-START=1 -END=1 -MATRIX=510sum62 -TRANS=human40.cdi -LIST=1 -DGCALIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=1 -MODB=LOCAL -OUTFMT=pfo
-NORM=ext -HEAPSIZES=500 -MINLEN=0 -MAXLEN=200000000 -NCPU=6 -NO_XLIFY
-NRG SCORES=0 -LONGLOG -THREADS=1 -GAPOF=10 -XGAPEXT=0.5 -FGAPOF=6 -FGAPEXT=7
-YGAPOF=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database : us-10-041-018-361:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	ID	Description
1	4113	100.0	2792	1	us-10-041-018-361	

ALIGNMENTS

RESULT 1

US-10-041-018-361

Alignment Scores:

Pred. No.:	0	Length:	2792
Score:	413.00	Matches:	784
Percent Similarity:	100.00	Conservative:	0
Best Local Similarity:	100.00	Mismatches:	0
Query Match:	100.00	Indels:	0
Ds:	1	Gaps:	0

us-10-041-018-383 (1-784) x us-10-041-018-361 (1-2792)

Ov
1 MetAsnLeuSerIeuCysTlleAlaserProIeuIenThIueGorXanZurDuo1131-30

Db
174 ATGAATCTTTCACATATGCATCGCGTCCCTTTGTTAACCAATCAATCGTCCCGGGCT 233

QY	21	LeuSerAlaIleHisThrAlaSerThrSerHisGlyGlnThrAsnProThrAsnLeu	40
DB	234	CTGTCAGCTATTTCATACAGCATCAACTTCACATGCTGGACAAACTAATCCCACTAATCTG	233
QY	41	IleIleAspThrThrIysGluArgIleGlnIysGlnPheLysAsnValGluIleSerVal	60
DB	294	ATCATTTGATACACAAAGAAGCGGATCCAAACAACAGTTTAAAAATGTAGAAATTTCTGTT	353
QY	61	SerSerTyrAspThrAlaTyrValAlaMetValProSerProAsnSerProLysSerPro	80
DB	354	TCTTCATATGACACAGCATGGGTAGCCATGGTCCCTTCTCCAAATCACCCAAATCGCCT	413
QY	81	CysPheProGluCysLeuAsnTyrPheLeuIleAsnAsnGlnLeuAsnAspGlySerTyrGly	100
DB	414	TGTTTCCCTGAGTGCTCAATTGGTTAATTAATATCAGCTTAATGATGGTTTCATGGGT	473
QY	101	LeuValAsnHisThrHisAsnHisAsnHisProLeuLeuLysAspSerLeuSerSerThr	120
DB	474	CTTGTTAATACACATCTCAATAATCAATAACCCGTTGCTTAAAGATTTCTATATCTCAACA	533
QY	121	LeuAlaCysIleValAlaLeuLysArgTyrAsnValGlyGluAspGlnIleAsnLysGly	140
DB	534	TTACCATGTATTGTTGCTTAAAGAAGATGGATGTTGGGGAAGATCAATAAATAAAGGT	593
QY	141	LeuSerPheIleGluSerAsnLeuAlaSerAlaThrGluLysSerGlnProSerProIle	160
DB	594	CTAAGTTTTATTGAGTCAAACTCTTGCTTCAGCTACTGAAAAAGTCAACCATCTCCCAT	653
QY	161	GlyPheAspIleIlePheProGlyLeuLeuGluTyrAlaLysAsnLeuAspIleAsnLeu	180
DB	654	GGTTTTGCACATCAATTTCTCTGGTTTGCTTGAGTATGCGAAAACTTGGACATAAACCTC	713
QY	181	LeuSerLysGlnThrAspPheSerLeuMetLeuHisLysArgGluLeuGluGlnLysArg	200
DB	714	CTTTCAAAAACAACAGATTTTAGTTTGATGCTACATAGAGGGAATTGGAGCAAAAAAGA	773
QY	201	CysHisSerAsnGluMetAspGlyTyrLeuAlaTyrIleSerGluGlyLeuGlyAsnLeu	220
DB	774	TGCGATTTCAAATGAGATGGATGGATCTTGCGGTATATCTCTGAAGGACTCGGTAAATT	833
QY	221	TyrAspTyrAsnMetValLysLysTyrGlnMetLysAsnGlySerValPheAsnSerPro	240
DB	834	TATGATTGGAATATGGTGAAGAAATATCAGATGAAAAATGGTTCTGTTTCAACTCACCA	893
QY	241	SerAlaThrAlaAlaPheIleAsnHisGlnAsnProGlyCysLeuAsnTyrLeuAsn	260
DB	894	TCAGCAACACGTGCTGCTTTCAITAAATCATCAAAATCTGTTGTCTTAATTTATTAAT	953
QY	261	SerLeuLeuAspLysPheGlyAsnAlaValProThrValTyrProHisAspLeuPheIle	280
DB	954	TCACTTTTGGACAAGTTTGTGTAATGTCAGTCCCAACAGTTTATCTCATGATTTATTTATC	1013
QY	281	ArgLeuSerMetValAspThrIleGluArgLeuGlyIleSerHisHisPheArgValGlu	300
DB	1014	CGACTTTCTATGTTGCACAAATGAAAGATTAGGAATTTACACACCATTTCCAGAGTGGA	1073
QY	301	IleLysAsnValLeuAspGluThrTyrArgCysTyrValGluArgAspGluGlnIlePhe	320
DB	1074	ATTTAAAAATGTTTTAGATGAAACATACAGATGTTGGGTGGAACAGAGATGAGCAAAATTC	1133
QY	321	MetAspValValThrCysAlaLeuAlaPheArgLeuLeuArgIleAsnGlyTyrGluVal	340
DB	1134	ATGGATGTTGTAACATGTGCTTTAGCCTTTCCGTTATTAGGATCAATGGGTATGAAGTT	1193
QY	341	SerProAspProLeuAlaGluIleThrAsnGluLeuAlaLeuLysAspGluTyrAlaAla	360
DB	1194	TCCCCAGATCCCATGGCTGAAAATTCAATGAATTAGCTTTTGAAGACGAATATGCAGCT	1253
QY	361	LeuGluThrTyrHisAlaSerHisIleLeuTyrGlnGluAspLeuSerSerGlyLysGln	380
DB	1254	CTTGAACAATATCATCGTACATATATATACCAAGAGATTTATCTCTGGAACAAACAA	1313

QY	381	IleLeuLysSerAlaaspPheLeuLysGluIleIleSerThrAspSerAsnArgLeuSer	400
Db	1314	ATCTTGAAGTCAGCTGATTTCTCAAAGAGATTAATATCCCTGATTCAAACAGGCTTCT	1373
QY	401	LysLeuIleHisLysGluValGluAsnAlaLeuLysPheProIleAsnThrGlyLeuGlu	420
Db	1374	AAATTAATTCACAAAGAGGTGGAAATGCTCTTAAGTTCCTTCAATACATCCGGTTAGAA	1433
QY	421	ArgIleAsnThrArgArgAsnIleGlnLeuLysPheAsnValAspAsnThrArgIleLeuLys	440
Db	1434	CGCATAAACACTAGACGAATATACAGCTTTCAATGTAGACAATCAAGAAATTCGAAA	1493
QY	441	ThrThrTyrHisSerSerAsnIleSerAsnThrAspTyrLeuArgLeuAlaValGluAsp	460
Db	1494	ACTACATATACATCATCAATATTAGTAACACTGANTACTTAAGTTCCTGCTGTGAAGAT	1553
QY	461	PheTyrThrCysGlnSerIleTyrArgGluGluLeuLysGlyLeuGluArgTyrValVal	480
Db	1554	TTCTCACACCTGCCAATCTATTATTCGTGAAGAAATTAAGAGTCTTGAAAGGTGGGTGTA	1613
QY	481	GluAsnLysLeuAspGlnLeuLysPheAlaArgGlnLysThrAlaTyrCysTyrPheSer	500
Db	1614	GAGAAATAAGTTGACACAGCTCAAGTTTGTAGCAAAAGACCGCCTACTGTATTCTCT	1673
QY	501	ValAlaAlaThrLeuSerSerProGluLeuSerAspAlaArgIleSerTrpAlaLysAsn	520
Db	1674	GTTCCTGCACACATTTTCGTCCTCCGAATATACAGATCGCGTATTTCAITGGGCCAAAAT	1733
QY	521	GlyIleLeuThrThrValValAspAspPhePheAspIleGlyGlyThrIleAspGluLeu	540
Db	1734	GGCATATTAACTACAGTAGTTGATGACTTTTGTGATATCGGTGGTACATCGATGAATTG	1793
QY	541	ThrAsnLeuIleGlnCysValGluLysTrpAsnValAspValAspLysAspCysSer	560
Db	1794	ACCAACCTGANTCAATGTGTGTAAGAAATGGAATGTAGATGTCACAGGATTTGTGTCA	1853
QY	561	GluHisValArgIleLeuPheLeuAlaLeuLysAspAlaIleCysTrpIleGlyAspGlu	580
Db	1854	GAGCATGTTCCGATTTTATTTTAGCATTTAAAGATGCATCTGTGGATGGAGATGAA	1913
QY	581	AlaPheLysTrpGlnAlaArgAspValThrSerHisValIleGlnThrTrpLeuGluLeu	600
Db	1914	GCATTTAAATGGCAAGCGCGCATGTAACTAGCCATCTTATTCAAACTTGGTTGGAAC	1973
QY	601	MetAsnSerMetLeuArgGluAlaIleTrpThrArgAspAlaTyrValProThrLeuAsn	620
Db	1974	ATGAATAGTATGTTGAGAAGAGCTATATGACACAGAGATGCTTATGTGCCAACATTAAT	2033
QY	621	GluTyrMetGluAsnAlaTyrValSerPheAlaLeuGlyProIleValLysProAlaIle	640
Db	2034	GAATATATGAAACCGCTTACGTGTCATTTGCTAGCCCGATTTGCAAGCCGGCTATT	2093
QY	641	TyrPheValGlyProLysLeuSerGluGluIleValGlnSerSerGluTyrHisAsnLeu	660
Db	2094	TACTTTGTGGGGCCCAATTTATCAGAGAGATTTGTGAAAGCTCTGAATATCATTAATCTA	2153
QY	661	PheLysLeuMetSerThrGlnGlyArgLeuLeuAsnAspIleHisSerPheLysArgGlu	680
Db	2154	TTTAAGCTAATAGACACGCGGGTCGACTTCTAAACGATATCCATAGCTTCAAGGGGAA	2213
QY	681	PheLysGluGlyLysLeuAsnAlaValAlaLeuHisLeuSerAsnGlyGluSerGlyLys	700
Db	2214	TTTAAGGAGGCAATTTAAACGCGGTAGCATTTGATTTGAGTAACGGAGAAAGTGGGAA	2273
QY	701	ValGluGluValValGluGluMetMetMetMetIleLysAsnLysArgLysGluLeu	720
Db	2274	GTGGAGAAGAGGTTGTGCGAGCATGATGATGATGATTAATAACAGAGAAAGAAATTA	2333
QY	721	MetLysLeuIlePheGluGluAsnGlySerIleValProArgAlaCysLysAspAlaPhe	740
Db	2334	ATGAATTAATTTTGAAGAAAATGTTAGCATTTCTCTAGAGCTTGTAAAGATGCATTT	2393
QY	741	TrpAsnMetCysHisValLeuAsnPhePheTyrAlaAsnAspGlyPheThrGlyAsn	760

Db	2394	TGGAACATGTGTACGTGTGTAATTTTTTTACGCAACCATGACGGGTTTACTGAAAC	2453
QY	761	ThrIleLeuAspThrValLysAspIleIleTyrAsnProLeuValLeuValAsnGluAsn	780
Db	2454	ACGATTCCTTGATCTGTGAGGACATCATTTACACCCGTTGGTGTGTGATGAAAT	2513
QY	781	GluGluGlnArg	784
Db	2514	GAAGAACAAAGG	2525

Search completed: October 18, 2004, 07:20:45
Job time : 8 secs

Db 241 TCRATTTTCAAGAGCTACTATAGAGAGAGCAAGAGTTTACGAGTCTGAAATCA 300
QY 301 ATGGAGGCCAAGATAGATGAGCTGATCAATAATGATCCCTGTTGGTGCAGCCAAATGAA 360
Db 301 ATGGAGGCCAAGATAGATGAGCTGATCAATAATGATCCCTGTTGGTGCAGCCAAATGAA 360
QY 361 AGCTTGATTTCAAAACCTTATATCAATCCCTTTGAAACCTGGCAGAACCTTTGACATA 420
Db 361 AGCTTGATTTCAAAACCTTATATCAATCCCTTTGAAACCTGGCAGAACCTTTGACATA 420
QY 421 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATAGTT 480
Db 421 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATAGTT 480
QY 481 TCGCAAAATTTGAGCTCTTGCAATAATTCAGCCTTTTAAATCGACATATAGAGATAAT 540
Db 481 TCGCAAAATTTGAGCTCTTGCAATAATTCAGCCTTTTAAATCGACATATAGAGATAAT 540
QY 541 GCTCCCTTGAGAGGGGACAGACCACTCTCACTTAACTTCGGTGTACCTCCACTATA 600
Db 541 GCTCCCTTGAGAGGGGACAGACCACTCTCACTTAACTTCGGTGTACCTCCACTATA 600
QY 601 AACACCGCAAAATTAATGATTTTTCAGAGCCATGCAACTTGTATCGCAGCTTAACCAAAA 660
Db 601 AACACCGCAAAATTAATGATTTTTCAGAGCCATGCAACTTGTATCGCAGCTTAACCAAAA 660
QY 661 GAGCCTTTGATCATAAATTTGATGATTTTCAAGAGAAATGATCAATCATATAGTAGG 720
Db 661 GAGCCTTTGATCATAAATTTGATGATTTTCAAGAGAAATGATCAATCATATAGTAGG 720
QY 721 GGCAGAGCTTGATATATATGATTTTTCAGAGCCATGCAACTTGTATCGCAGCTTAACCAAAA 780
Db 721 GGCAGAGCTTGATATATATGATTTTTCAGAGCCATGCAACTTGTATCGCAGCTTAACCAAAA 780
QY 781 ATGTAATTTGAATATGTTTATGATTAATAAAGAGGGGCTTTTTCAGATTAACCTTGAGACT 840
Db 781 ATGTAATTTGAATATGTTTATGATTAATAAAGAGGGGCTTTTTCAGATTAACCTTGAGACT 840
QY 841 ATGGAAGCGTGCTCTCTCTCTCTCAACAGCGGCAATTCGTTGGTCTTTTCAATAATCTT 900
Db 841 ATGGAAGCGTGCTCTCTCTCTCTCAACAGCGGCAATTCGTTGGTCTTTTCAATAATCTT 900
QY 901 CTGGTATTTATTTATCAGATTTAGATGATTTTGAATTTGAAAGATTTCCAAATCTCC 960
Db 901 CTGGTATTTATTTATCAGATTTAGATGATTTTGAATTTGAAAGATTTCCAAATCTCC 960
QY 961 AGCGAAAAGGCTTTGCTGAGGACATTTACAGAGGGGAGTTATCTTTTCCCATCGTCCAC 1020
Db 961 AGCGAAAAGGCTTTGCTGAGGACATTTACAGAGGGGAGTTATCTTTTCCCATCGTCCAC 1020
QY 1021 GCCCTTAACCTTCAATAAAAGGTCACCAAGGTCACCACTGAGCAACATGAAATCTAAGAA 1080
Db 1021 GCCCTTAACCTTCAATAAAAGGTCACCAAGGTCACCACTGAGCAACATGAAATCTAAGAA 1080
QY 1081 CTCCTGTTGAGGACAGGTATATAAGATATAAACTAAAGCTGATTTCAAAATCTGGAATTC 1140
Db 1081 CTCCTGTTGAGGACAGGTATATAAGATATAAACTAAAGCTGATTTCAAAATCTGGAATTC 1140
QY 1141 GACACCAATTCATTTGGCCCTACACCAAAATTTTAAATCAATTAATGATGATATGATAAA 1200
Db 1141 GACACCAATTCATTTGGCCCTACACCAAAATTTTAAATCAATTAATGATGATATGATAAA 1200
QY 1201 AATGATTAATGAAAATAGTATTTTACCTGATTTGGCTTGGCAATTCGACACCGCCACCAAT 1260
Db 1201 AATGATTAATGAAAATAGTATTTTACCTGATTTGGCTTGGCAATTCGACACCGCCACCAAT 1260
QY 1261 TTACATGACGAATTTGTTATATATATAGACCACTTATCCGAAATTTGTAATTAATGATC 1320
Db 1261 TTACATGACGAATTTGTTATATATATAGACCACTTATCCGAAATTTGTAATTAATGATC 1320
QY 1321 AATCAAAATTTAGTGGAGGAGATAGTCAGAAATTAAGAGCTTCTCTCTCTCTCTCTCTCT 1380

1321 AATCAAAATTTAGTGGAGGAGATAGTCAGAAATTAAGAGCTTCTCTCTCTCTCTCTCTCT 1380
QY 1381 TATACATAGCAATTTTATATATATACGTTTTCATTCGATCATCTTTTATGATATATCTCAAAA 1440
Db 1381 TATACATAGCAATTTTATATATATACGTTTTCATTCGATCATCTTTTATGATATATCTCAAAA 1440
QY 1441 TCTCTTAGTTTGCAGAAATAGTCAAAATCTTCAAAATTTATAGCTTTTATATTTTCCACGAT 1500
Db 1441 TCTCTTAGTTTGCAGAAATAGTCAAAATCTTCAAAATTTATAGCTTTTATATTTTCCACGAT 1500
QY 1501 TTCTGAAATCTCTTTTATATCAGACCGTTAAATGCTAGCGTTTACTGTCAAAATCGCCGTA 1560
Db 1501 TTCTGAAATCTCTTTTATATCAGACCGTTAAATGCTAGCGTTTACTGTCAAAATCGCCGTA 1560
QY 1561 AANTGCGGA 1569
Db 1561 AANTGCGGA 1569

RESULT 2
SCU31632
LOCUS
DEFINITION
SCU31632
1569 bp DNA linear PLN 14-DBC-1995
Saccharomyces cerevisiae geranylgeranyl diphosphate synthase (BTS1)
gene, complete cds.
U31632.1 GI:1098640
SOURCE
Saccharomyces cerevisiae (baker's yeast)
ORGANISM
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
1 (bases 1 to 1569)
Jiang, Y., Proteau, P., Poulter, D. and Ferro-Novick, S.
BTS1 encodes a geranylgeranyl diphosphate synthase in Saccharomycetes
cerevisiae
J. Biol. Chem. 270 (37), 21793-21799 (1995)
J. Biol. Chem. 270 (37), 21793-21799 (1995)
MEDLINE
95394944
PUBMED
765600
REFERENCE
2 (bases 1 to 1569)
Jiang, Y. and Ferro-Novick, S.
Direct Submission
Submitted (14-JUL-1995) Yu Jiang, Cell Biology, Yale University,
295 Congress Ave., New Haven, CT 06510, USA
JOURNAL
Location/Qualifiers
FEATURES
1..1569
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
301..1308
/gene="BTS1"
301..1308
/gene="BTS1"
/notes="geranylgeranyl diphosphate synthase"
/codon_start=1
/product="Bts1p"
/protein_id="AAA83262.1"
/db_xref="GI:1098641"
/translation="MEAKIDELINNDPWSQNESLTSKYPNHLILPKGNFRNLIV
QINVMNLPDQLAIYSQIVVELLHNSLLIDTENDAPLRGOTTSHLIFGPSTINT
ANTYFRAMQVLSQLTKEPLYENLITFNEELINLHRGGLDIYWRDLPEIPTQE
MYLNVMNKVQGLFRILRLMBEALSPSSHGHSLVFPINLLGIIYQIRDDYLNLDKQ
MSSEKPAEDITEGKLSFPITVHALNFTKTGQEQHNEILRLILRLSDKIDKLIQ
ILSFDTNSLAYTKNFINQLVNMINKDNENKYLPLDASHSDTAINLHDELLIYIDHLE
L"

Query Match 100.0%; Score 1569; DB 8; Length 1569;
Best Local Similarity 100.0%; Pred. No. 3.7e-309; Indels 0; Gaps 0;
Matches 1569; Conservative 0; Mismatches 0;
QY 1 AATATTATATATAGATAGGACCAAGCCGCAATTTTATATCTGAAAGTAAACTTCTATT 60
Db 1 AATATTATATATAGATAGGACCAAGCCGCAATTTTATATCTGAAAGTAAACTTCTATT 60

QY 61 ATTATAGTGGTATCCAAAGCTTCCACCGCTTCCAGCATAGCAGAAATACGTGTTTTCGAT 120
Db 61 ATTATAGTGGTATCCAAAGCTTCCACCGCTTCCAGCATAGCAGAAATACGTGTTTTCGAT 120
QY 121 ATGTTATGCTGATCATTTGATGCTTACTACCAATTTTCTTTCCTTCGCGCTTCCTTCTTT 180
Db 121 ATGTTATGCTGATCATTTGATGCTTACTACCAATTTTCTTTCCTTCGCGCTTCCTTCTTT 180
QY 181 GACGTTTTTTTGAACCAAAAAAAGTCAAGCAGATGCTTCAAAAACCATGTAAGGC 240
Db 181 GACGTTTTTTTGAACCAAAAAAAGTCAAGCAGATGCTTCAAAAACCATGTAAGGC 240
QY 241 TCAATTTTCAAAGAGCTACTAATAAGAGAGAAAGATTTTACAGTCTGCGAAATCA 300
Db 241 TCAATTTTCAAAGAGCTACTAATAAGAGAGAAAGATTTTACAGTCTGCGAAATCA 300
QY 301 ATGGAGGCCAAGATAGATGAGCTGATCAATATGATCTCTGTTTGGTCCAGCCAAATGAA 360
Db 301 ATGGAGGCCAAGATAGATGAGCTGATCAATATGATCTCTGTTTGGTCCAGCCAAATGAA 360
QY 361 AGCTTGATTTCAAACCTTATAATCAGATCTCTTTTGAACCTGGCAAGAACTTTAGACTA 420
Db 361 AGCTTGATTTCAAACCTTATAATCAGATCTCTTTTGAACCTGGCAAGAACTTTAGACTA 420
QY 421 AATTTAATAGTTCAAATTAACAGAGTATGATTTTGCACCAAGACAGCTGSCCATAGTT 480
Db 421 AATTTAATAGTTCAAATTAACAGAGTATGATTTTGCACCAAGACAGCTGSCCATAGTT 480
QY 481 TCGCAATATGTTGAGCTTTGATATTAATCCAGCTTTTAAATCGACGATATAGAGATAAT 540
Db 481 TCGCAATATGTTGAGCTTTGATATTAATCCAGCTTTTAAATCGACGATATAGAGATAAT 540
QY 541 GCTCCTTGAGAGGGGACAGACACTTCTCACTTAATCTTCGTTGTTACCTTCCACTATA 600
Db 541 GCTCCTTGAGAGGGGACAGACACTTCTCACTTAATCTTCGTTGTTACCTTCCACTATA 600
QY 601 AACACCGCAATATATGTTTTCAGAGCCATGCAACTTGTATGCGACGTAACCAACAAAA 660
Db 601 AACACCGCAATATATGTTTTCAGAGCCATGCAACTTGTATGCGACGTAACCAACAAAA 660
QY 661 GAGCTTTGATATCAATTTGATACGATTTTCAACGAGAAATGATCAATCTACATAGG 720
Db 661 GAGCTTTGATATCAATTTGATACGATTTTCAACGAGAAATGATCAATCTACATAGG 720
QY 721 GGACAGGCTTGGATATATCTGAGAGACTTTCGCTGGAATCATACCTACTCAGAG 780
Db 721 GGACAGGCTTGGATATATCTGAGAGACTTTCGCTGGAATCATACCTACTCAGAG 780
QY 781 ATGTATTGGAATATGTTTATGAATAAAACAGCGGCTTTTTCAGATTAAAGTTGAGACTC 840
Db 781 ATGTATTGGAATATGTTTATGAATAAAACAGCGGCTTTTTCAGATTAAAGTTGAGACTC 840
QY 841 ATGGAAGCGCTGTCTCTCTCCACACACGCGCATTCGTTGGTTCCCTTTCATAAATCTT 900
Db 841 ATGGAAGCGCTGTCTCTCTCCACACACGCGCATTCGTTGGTTCCCTTTCATAAATCTT 900
QY 901 CTGGGTATTATTATCAGATTAGAGATGATTAATTTGAAGATTTCCAAATGTC 960
Db 901 CTGGGTATTATTATCAGATTAGAGATGATTAATTTGAAGATTTCCAAATGTC 960
QY 961 AGCGAAAAAGCTTGTCTGAGGACATTAACAGAGGAGATTTCTTTTCCCATCGTCCAC 1020
Db 961 AGCGAAAAAGCTTGTCTGAGGACATTAACAGAGGAGATTTCTTTTCCCATCGTCCAC 1020
QY 1021 GCCCTTAACTTCAATAAAAGGTCAACTGAGCAACAATGAATGAATTTCTAAGAAAT 1080
Db 1021 GCCCTTAACTTCAATAAAAGGTCAACTGAGCAACAATGAATGAATTTCTAAGAAAT 1080
QY 1081 CTCCTGTTGAGGACAGTGATAAAGATATAAATCAAGCTGATTTCAAATCTGAATTC 1140
Db 1081 CTCCTGTTGAGGACAGTGATAAAGATATAAATCAAGCTGATTTCAAATCTGAATTC 1140

QY 1141 GACACCAATTCATTGGCTCAGACCAAAAAATTTTATTATCAATTAATGATATGATATAA 1200
Db 1141 GACACCAATTCATTGGCTCAGACCAAAAAATTTTATTATCAATTAATGATATGATATAA 1200
QY 1201 AATGATATGAAAAATAGTATTACCTGATTTGGCTTCGCATTCGCACACGCCCAACAT 1260
Db 1201 AATGATATGAAAAATAGTATTACCTGATTTGGCTTCGCATTCGCACACGCCCAACAT 1260
QY 1261 TTACATGACGAATTTGATATATATATAGACCACTTATCCGAATTCGAAATATAATGATC 1320
Db 1261 TTACATGACGAATTTGATATATATATAGACCACTTATCCGAATTCGAAATATAATGATC 1320
QY 1321 AATCAAAATAGTGGGAGAGTAGTCAGAAATAAGCCCTTCTCTCTCTCTCTCTCTCT 1380
Db 1321 AATCAAAATAGTGGGAGAGTAGTCAGAAATAAGCCCTTCTCTCTCTCTCTCTCTCT 1380
QY 1381 TATACATACGAATTTTAT 1440
Db 1381 TATACATACGAATTTTAT 1440
QY 1441 TCTCTTAGTTCGCAAAATAGTCAAAATCTTCAAAATTTATAGCCCTTTATATATTTTCCACGAT 1500
Db 1441 TCTCTTAGTTCGCAAAATAGTCAAAATCTTCAAAATTTATAGCCCTTTATATATTTTCCACGAT 1500
QY 1501 TTCTGAACTCTCTTTTATATAGCACCCGTTAATGCTAGCGGTACTGTCAAAATCGCGGTA 1560
Db 1501 TTCTGAACTCTCTTTTATATAGCACCCGTTAATGCTAGCGGTACTGTCAAAATCGCGGTA 1560
QY 1561 AATTGCGGA 1569
Db 1561 AATTGCGGA 1569

RESULT 3
SCU39205/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REMARK
REFERENCE
AUTHORS

SCU39205 37947 bp DNA linear PLN 02-SEP-1997
Saccharomyces cerevisiae chromosome XVI, left arm, cosmid 8460.
U39205 U00094
U39205.1 GI:1079672
geranylgeranyl pyrophosphate homolog; aldehyde dehydrogenase
homolog; S. cerevisiae Mrs2p homolog; rRNA-Cys; Ty4;
glutaredoxin-like homolog; S. cerevisiae Pdr5p homolog; SURI; LEE1;
KTR6; D. melanogaster ARL2 homolog.
Saccharomyces cerevisiae (baker's yeast)
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
1 (bases 29075 to 30223)
Defarges, L., Durrens, P., Juguelin, H., Cassagne, C., Bonneau, M. and
Aigle, M.
Yeast mutants affected in viability upon starvation have a modified
phospholipid composition
Yeast 9 (3), 267-277 (1993)
93255905
8488727
SUP1 gene designation from this publication.
2 (bases 1 to 37947)
Bussey, H., Storms, R.K., Ahmed, A., Albermann, K., Allen, E.,
Ansoy, W., Araujo, R., Aparicio, A., Barrell, B., Badcock, K.,
Benes, V., Botstein, D., Bowman, S., Bruckner, M., Carpenter, J.,
Cherry, J.M., Chung, E., Church, C., Coster, F., Davis, K.,
Davis, R.W., Dietrich, F.S., Dilius, H., Dipaolo, T., Dubois, E.,
Dusterhoft, A., Duncan, M., Floeth, M., Fortin, N., Friesen, J.D.,
Fritz, C., Goffeau, A., Hall, J., Hebling, U., Heumann, K., Hilbert, H.,
Hillier, L., Hunnicke-Smith, S., Hyman, R., Johnston, M., Kalman, S.,
Kleine, K., Komp, C., Kurdi, O., Lashkari, D., Lew, H., Lin, A., Lin, D.,
Louis, E.J., Marathe, R., Messenguy, F., Mewes, H.W., Mirtipati, S.,
Moestl, D., Muller-Auer, S., Namath, A., Nentwich, U., Oefner, P.,
Pearson, D., Pecci, F., Pohl, T.M., Purnelle, D., Schafer, M.,
Scharfe, M., Scherens, B., Schram, S., Schroeder, M., Sdicu, A.M.,
Tettelin, H., Urrestarazu, L.A., Ushinsky, S., Vierendeels, F.,
Vissers, S., Voss, H., Walsh, S.V., Wambutt, R., Wang, Y., Wedler, E.,
Wedler, H., Winnett, E., Zhong, W.W., Zollner, A., Vo, D.H. and Ham, J.

CDS	/gene="YPL062W" 9061..9465 /gene="YPL062W" /codon_start=1 /product="Ypl062wp" /protein_id="AAB68303.1" /db_xref="GI:1079680" /translation="MTELQYVKGEDTIVETATSTPMLRSPILARAEEKRGLFPDIN EGHGHRDIVAVGVSKPKRSPHPHGKAAADKKRTTEPERKKRVGAPRCKKICI OLLRYRNFVLGIGNVCNIDLLGLTVSAINA" 9754..11256 /gene="ALD6" 9754..11256 /gene="ALD6" /notes="YPL061W" /codon_start=1	
gene		
CDS		
Query Match	86.4%; Score 1356.4; DB 8; Length 37947;	
Best Local Similarity	99.9%; Pred. No. 4.3e-266;	
Matches 1357; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 AATATTACATATAGATATAGGACAGCCGCGCATTTTTCATACCTCAAGGTAAACTTCTATT	60
DB	1358 AATATTACATATAGATATAGGACAGCCGCGCATTTTTCATACCTCAAGGTAAACTTCTATT	1299
QY	61 ATTATAGTGGTATCCAAAGCTTCCAGCGTTCCAGCATAGCAGAAATTACGCTGTTTGTGCAT	120
DB	1298 ATTATAGTGGTATCCAAAGCTTCCAGCGTTCCAGCATAGCAGAAATTACGCTGTTTGTGCAT	1239
QY	121 ATGTATGCTGATCATTTGATGCTTACTACCAATTTTCTTGTGCTTGCCTTCTTT	180
DB	1238 ATGTATGCTGATCATTTGATGCTTACTACCAATTTTCTTGTGCTTGCCTTCTTT	1179
QY	181 GACGTTTTTTTCAAGCAAAAAAAGTCAAGACAGATGCTTACAAAACGATGTAAGGC	240
DB	1178 GACGTTTTTTTCAAGCAAAAAAAGTCAAGACAGATGCTTACAAAACGATGTAAGGC	1119
QY	241 TCATTTTCAAAGAGCTACTAATAGAAAAGAGAAAGATTTACGAGTCTCGAAAAATCA	300
DB	1118 TCATTTTCAAAGAGCTACTAATAGAAAAGAGAAAGATTTACGAGTCTCGAAAAATCA	1059
QY	301 ATGGAGGCAAGATAGATGAGTGATCAATAATGATCCTGTTGTCAGGCCAAATGAA	360
DB	1058 ATGGAGGCAAGATAGATGAGTGATCAATAATGATCCTGTTGTCAGGCCAAATGAA	999
QY	361 AGCTTGATTTCAAACCTTATATACATCCTTTTGAACCTGCGAAGACTTTAGACTA	420
DB	998 AGCTTGATTTCAAACCTTATATACATCCTTTTGAACCTGCGAAGACTTTAGACTA	939
QY	421 AATTTAATGTTCAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATAGTT	480
DB	938 AATTTAATGTTCAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATAGTT	879
QY	481 TCGCAATTTGAGCTCTTGCATTAATTCAGAGCTTTTAAATCGACATATAGAGATAAT	540
DB	878 TCGCAATTTGAGCTCTTGCATTAATTCAGAGCTTTTAAATCGACATATAGAGATAAT	819
QY	541 GTCCTTTGAGAGGGGACAGACCTTCTCAGTTAATCTTGGTGTAACCTTCCACTATA	600
DB	818 GTCCTTTGAGAGGGGACAGACCTTCTCAGTTAATCTTGGTGTAACCTTCCACTATA	759
QY	601 AACACCGCAATTAATGATTTTTCAGAGCCATGCAATTTGATCGACCTAACCCACAAAA	660
DB	758 AACACCGCAATTAATGATTTTTCAGAGCCATGCAATTTGATCGACCTAACCCACAAAA	699
QY	661 GAGCTTTGATATCAATTTGATGATGATTTTCAACGAGAAATGATCAATCTACATAGG	720
DB	698 GAGCTTTGATATCAATTTGATGATGATTTTCAACGAGAAATGATCAATCTACATAGG	639
QY	721 GGACAGGCTTGGATATATATCTGAGAGACTTCTGCTGCTGAATCATACCTACTCAGAG	780
DB	638 GGACAGGCTTGGATATATATCTGAGAGACTTCTGCTGCTGAATCATACCTACTCAGAG	579
QY	781 AUGTATTGAATATGCTTATGATATAAACAAGCGGCGCTTTTCAGATTACGTTGAGACTC	840
DB	578 AUGTATTGAATATGCTTATGATATAAACAAGCGGCGCTTTTCAGATTACGTTGAGACTC	519
QY	841 ATGGAAGCGCTGCTCTCTTCTCACACACGCGGCATTTGCTTGGTCTCTTTTCATAAATCTT	900
DB	518 ATGGAAGCGCTGCTCTCTTCTCACACACGCGGCATTTGCTTGGTCTCTTTTCATAAATCTT	459
QY	901 CTGGGTATTATTATTCAGATTAGAGATGATTTACTTGAATTTGAAAGATTTCCAAATGTCC	960
DB	458 CTGGGTATTATTATTCAGATTAGAGATGATTTACTTGAATTTGAAAGATTTCCAAATGTCC	399
QY	961 AGCGAAAAGGCTTTGCTGAGGACATTTACAGAGGGGAAGTTATCTTTTCCATCGTCCAC	1020
DB	398 AGCGAAAAGGCTTTGCTGAGGACATTTACAGAGGGGAAGTTATCTTTTCCATCGTCCAC	339
QY	1021 GCCCTTTAACTTCACTAAAAAGGCTCAAACTGAGCAACACAAATGAAATTTCTAAGAAAT	1080
DB	338 GCCCTTTAACTTCACTAAAAAGGCTCAAACTGAGCAACACAAATGAAATTTCTAAGAAAT	279
QY	1081 CTCCTGTTGAGGACAGTGATATAAGATATAAACTAAAGCTGATTCAAATCTGGAATTC	1140
DB	278 CTCCTGTTGAGGACAGTGATATAAGATATAAACTAAAGCTGATTCAAATCTGGAATTC	219
QY	1141 GACACCAATTCATTCGCCCTTACACCAAAAATTTTATTAATCAATTAGTGAATATGATAAA	1200
DB	218 GACACCAATTCATTCGCCCTTACACCAAAAATTTTATTAATCAATTAGTGAATATGATAAA	159
QY	1201 AATGATAATGAAAAAAGTATTACCTGATTTGGCTTCGCATTCGACACGCCACCAAT	1260
DB	158 AATGATAATGAAAAAAGTATTACCTGATTTGGCTTCGCATTCGACACGCCACCAAT	99
QY	1261 TTACATGACGAATGTTTATATATAATAGACCACTTTATCCGAATTTGAAAAATTAATGATC	1320
DB	98 TTACATGACGAATGTTTATATATAATAGACCACTTTATCCGAATTTGAAAAATTAATGATC	39
QY	1321 AATCAAAATTTAGTGGAGGAGATAGTCAGAAATAAGCC	1358
DB	38 AATCAAAATTTAGTGGAGGAGATAGTCAGAAATAAGCC	1
RESULT 4		
BD177419	1008 bp DNA linear	PAT 16-APR-2003
LOCUS	A method of producing prenyl alcohol.	
DEFINITION	BD177419	
ACCESSION	BD177419.1	GI:30014680
VERSION	JP 2002300896-A/12.	
KEYWORDS	Saccharomyces cerevisiae (baker's yeast)	
SOURCE	Saccharomyces cerevisiae	
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.	
REFERENCE	1 (bases 1 to 1008)	
AUTHORS	Muramatsu, M., Kobata, M. and Shimizu, A.	
TITLE	A method of producing prenyl alcohol	
JOURNAL	Patent: JP 2002300896-A 12 15-OCT-2002;	
COMMENT	TOYOTA MOTOR CORP	
OS	Saccharomyces cerevisiae (yeast)	
PN	JP 2002300896-A/12	
PD	15-OCT-2002	
PF	18-JAN-2002	JP 2002010528
PI	MASAYOSHI MURAMATSU, MITSUO KOBATA, AKIRA SHIMIZU	PC
CL2P7/04	1 (12P7/04, 12R1:85), (12P7/04, 12R1:78), (12P7/04, 12R1:84),	
PC	(12P7/04, 12R1:72), (12P7/04, 12R1:44), (12P7/04, 12R1:265),	
PC	(12P7/04, 12R1:05), (12P7/04, 12R1:01), (12P7/04, 12R1:645)	
CC	A method of producing prenyl alcohol	
FT	Key Location/Qualifiers	
FT	CDS (1) . (1005).	
FEATURES	1..1008	
source	/organism="Saccharomyces cerevisiae"	
	/mol_type="genomic DNA"	

Db		961 TTACATGACGAATTGTTATATATATATATATATAGACCACTTATCGGAATTGTGA 1008	
ORIGIN			
Query Match		64.2%; Score 1008; DB 6; Length 1008;	
Best Local Similarity		100.0%; Pred. No. 5.9e-195;	
Matches 1008; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
QY	301	ATGGAGCCCAAGATAGATGAGCTGATCAATATGATCTGTTGGTCCAGCCAAATGAA 360	
Db	1	ATGGAGCCCAAGATAGATGAGCTGATCAATATGATCTGTTGGTCCAGCCAAATGAA 60	
QY	361	AGCTTGATTTCAAACCTTATATACATCTTTGAAACCTGCGAAGAACTTTAGACTA 420	
Db	61	AGCTTGATTTCAAACCTTATATACATCTTTGAAACCTGCGAAGAACTTTAGACTA 120	
QY	421	AATTTAATAGTTCAAATTAACAGAGTTATGAATTTGCCAAAGACCGAGTGGCCATGTT 480	
Db	121	AATTTAATAGTTCAAATTAACAGAGTTATGAATTTGCCAAAGACCGAGTGGCCATGTT 180	
QY	481	TGCAAAATTTGTTGAGCTCTTGCAATAATTCAGCCCTTTTAATTCGACGATATAGAGATAAT 540	
Db	181	TGCAAAATTTGTTGAGCTCTTGCAATAATTCAGCCCTTTTAATTCGACGATATAGAGATAAT 240	
QY	541	GCTCCCTTGAGAGGGGACAGCCACTTCTCACTTAATCTTCGGTGTACCCCTCCACTATA 600	
Db	241	GCTCCCTTGAGAGGGGACAGCCACTTCTCACTTAATCTTCGGTGTACCCCTCCACTATA 300	
QY	601	AACACCCCAAATTTATATGTTTTCAGAGCCATGCAATTTGATTCGAGCTTAACCAAAA 660	
Db	301	AACACCCCAAATTTATATGTTTTCAGAGCCATGCAATTTGATTCGAGCTTAACCAAAA 360	
QY	661	GAGCTTTGATCATATATTTGATTACGATTTTCAACGAAGAAATTTGATCAATCTACATAGG 720	
Db	361	GAGCTTTGATCATATATTTGATTACGATTTTCAACGAAGAAATTTGATCAATCTACATAGG 420	
QY	721	GGACAAAGCTTTGGATATATATCTGGAGAGACTTTTCGCTGAAATCATACCTACTCAGGAG 780	
Db	421	GGACAAAGCTTTGGATATATATCTGGAGAGACTTTTCGCTGAAATCATACCTACTCAGGAG 480	
RESULT 5			
BD170961			
LOCUS		1008 bp DNA linear PAT 17-JAN-2003	
DEFINITION		Process for producing prenyl alcohol.	
ACCESSION		BD170961	
VERSION		BD170961.1 GI:27876773	
KEYWORDS		WO 02053746-A/3.	
SOURCE		Saccharomyces cerevisiae (baker's yeast)	
ORGANISM			
REFERENCE			
AUTHORS		Oto, C., Obata, S., Muramatsu, M., Nishi, K. and Totsumura, K.	
TITLE		Process for producing prenyl alcohol	
JOURNAL		Patent: WO 02053746-A 3 11-JUL-2002;	
COMMENT		TOYOTA MOTOR CORP, CHIKARA OTO, SHUSEI OBATA, MASAYOSHI MURAMATSU, KIYOHICO NISHI, KAZUHIKO TOTSUKA	
FEATURES			
source		1.1008	
ORIGIN		/organism="Saccharomyces cerevisiae"	
Query Match		64.2%; Score 1008; DB 6; Length 1008;	
Best Local Similarity		100.0%; Pred. No. 5.9e-195;	
Matches 1008; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
QY	301	ATGGAGCCCAAGATAGATGAGCTGATCAATATGATCTGTTGGTCCAGCCAAATGAA 360	
Db	1	ATGGAGCCCAAGATAGATGAGCTGATCAATATGATCTGTTGGTCCAGCCAAATGAA 60	
QY	361	AGCTTGATTTCAAACCTTATATACATCTTTGAAACCTGCGAAGAACTTTAGACTA 420	
Db	61	AGCTTGATTTCAAACCTTATATACATCTTTGAAACCTGCGAAGAACTTTAGACTA 120	
QY	421	AATTTAATAGTTCAAATTAACAGAGTTATGAATTTGCCAAAGACCGAGTGGCCATGTT 480	
Db	121	AATTTAATAGTTCAAATTAACAGAGTTATGAATTTGCCAAAGACCGAGTGGCCATGTT 180	
QY	481	TGCAAAATTTGTTGAGCTCTTGCAATAATTCAGCCCTTTTAATTCGACGATATAGAGATAAT 540	
Db	181	TGCAAAATTTGTTGAGCTCTTGCAATAATTCAGCCCTTTTAATTCGACGATATAGAGATAAT 240	
QY	541	GCTCCCTTGAGAGGGGACAGCCACTTCTCACTTAATCTTCGGTGTACCCCTCCACTATA 600	
Db	241	GCTCCCTTGAGAGGGGACAGCCACTTCTCACTTAATCTTCGGTGTACCCCTCCACTATA 300	
QY	601	AACACCCCAAATTTATATGTTTTCAGAGCCATGCAATTTGATTCGAGCTTAACCAAAA 660	
Db	301	AACACCCCAAATTTATATGTTTTCAGAGCCATGCAATTTGATTCGAGCTTAACCAAAA 360	
QY	661	GAGCTTTGATCATATATTTGATTACGATTTTCAACGAAGAAATTTGATCAATCTACATAGG 720	
Db	361	GAGCTTTGATCATATATTTGATTACGATTTTCAACGAAGAAATTTGATCAATCTACATAGG 420	
QY	721	GGACAAAGCTTTGGATATATATCTGGAGAGACTTTTCGCTGAAATCATACCTACTCAGGAG 780	
Db	421	GGACAAAGCTTTGGATATATATCTGGAGAGACTTTTCGCTGAAATCATACCTACTCAGGAG 480	

```
QY 781 ATGTATTGAATATGTTATGAATAAAGAGGGGCTTTTTCAGATTAACGTTGAGACTC 840
|
|
|
Db 481 ATGTATTGAATATGTTATGAATAAAGAGGGGCTTTTTCAGATTAACGTTGAGACTC 540
|
|
|
QY 841 ATGGAAGCGCTGTCCTTCCTCACAACGCGGCATTCGTTGGTTCCTTTCATAAATCTT 900
|
|
|
Db 541 ATGGAAGCGCTGTCCTTCCTCACAACGCGGCATTCGTTGGTTCCTTTCATAAATCTT 600
|
|
|
QY 901 CTGGGTATTATTATCAGATTAGAGATGATTACTTCAATTTGAAAGATTTCAAATGTCC 960
|
|
|
Db 601 CTGGGTATTATTATCAGATTAGAGATGATTACTTCAATTTGAAAGATTTCAAATGTCC 660
|
|
|
QY 961 AGCGAAAAAGGCTTGTCTGAGGACATTAAGAGGGGAGTATCTTTTCCCATCGTCCAC 1020
|
|
|
Db 661 AGCGAAAAAGGCTTGTCTGAGGACATTAAGAGGGGAGTATCTTTTCCCATCGTCCAC 720
|
|
|
QY 1021 GCCCTTAATCTTCACTAAAGAGAAAGGTCAAACTGAGCAACAATGAAATCTTAAGAAT 1080
|
|
|
Db 721 GCCCTTAATCTTCACTAAAGAGAAAGGTCAAACTGAGCAACAATGAAATCTTAAGAAT 780
|
|
|
QY 1081 CTCCTGTTGAGGACAAAGTATAAAGATATAAAGTAAAGCTGATTCAAATAGTGAATTA 1140
|
|
|
Db 781 CTCCTGTTGAGGACAAAGTATAAAGATATAAAGTAAAGCTGATTCAAATAGTGAATTA 840
|
|
|
QY 1141 GACACCAATTCATTTGGCTTACACCAAAAATTTTATTAATCAATTAAGTGAATATAAATA 1200
|
|
|
Db 841 GACACCAATTCATTTGGCTTACACCAAAAATTTTATTAATCAATTAAGTGAATATAAATA 900
|
|
|
QY 1201 AATGATATGAATAAGTATTTACTGATTTGGCTTCGATTCGACACCGCCACCAAT 1260
|
|
|
Db 901 AATGATATGAATAAGTATTTACTGATTTGGCTTCGATTCGACACCGCCACCAAT 960
|
|
|
QY 1261 TTACATGACGAATTTGTTATATATATAGACCACTTATCCGAATGTGA 1308
|
|
|
Db 961 TTACATGACGAATTTGTTATATATATAGACCACTTATCCGAATGTGA 1008
|
|
|
RESULT 6
LOCUS BD171079 1008 bp DNA linear PAT 17-JAN-2003
DEFINITION Process for producing prenyl alcohol.
ACCESSION BD171079
VERSION BD171079.1 GI:27876891
KEYWORDS WO 02053747-A/3.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
1 (bases 1 to 1008)
Oto,C. and Obata,S.
Process for producing prenyl alcohol
Patent: WO 02053747-A 3 11-JUL-2002;
TOYOFA MOTOR CORP,CHIKARA OTO,SHUSEI OBATA
OS Saccharomyces cerevisiae (yeast)
PN WO 02053747-A/3
PD 11-JUL-2002
PF 20-DEC-2001 WO 2001JP011215
PR 28-DEC-2000 JP 00P 401701.28-DEC-2000 JP 00P 403067 PR
18-SEP-2001 JP 01P 282978
PI CHIKARA OTO,SHUSEI OBATA
PC C12N15/52,C12P7/04,C12N1/19,C12N1/21
CC Process for producing prenyl alcohol
FH Key Location/Qualifiers
FT CDS (1)..(1005).
FEATURES
source
1..1008
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
ORIGIN
Query Match 64.2%; Score 1008; DB 6; Length 1008;
```

```
Best Local Similarity 100.0%; Pred. No. 5.9e-195;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 301 ATGGAGGCCAAGATAGATGAGCTGATCAATAATGATCTCTGTGTTGGTCCAGCCAAATGAA 360
|
|
|
Db 1 ATGGAGGCCAAGATAGATGAGCTGATCAATAATGATCTCTGTGTTGGTCCAGCCAAATGAA 60
|
|
|
QY 361 AGCTTGATTTCAAAACCTTAAATCAGATCCCTTTTGAACCTGGCAAGAACTTTAGACTA 420
|
|
|
Db 61 AGCTTGATTTCAAAACCTTAAATCAGATCCCTTTTGAACCTGGCAAGAACTTTAGACTA 120
|
|
|
QY 421 AATTTAAATAGTTCAAAATTAACAGAGTTATGAAATTTGCCCAAGACAGCTGCCATAGTT 480
|
|
|
Db 121 AATTTAAATAGTTCAAAATTAACAGAGTTATGAAATTTGCCCAAGACAGCTGCCATAGTT 180
|
|
|
QY 481 TCGCAAAATTTGTGAGCTCTTGCATAATTCAGAGCTTTTAAATCGACGATATAGAGATTAAT 540
|
|
|
Db 181 TCGCAAAATTTGTGAGCTCTTGCATAATTCAGAGCTTTTAAATCGACGATATAGAGATTAAT 240
|
|
|
QY 541 GTCCTCTTGAAGGGGACAGACCACTTCTCACTTAATCTTGGGTGTACCTTCCACTATA 600
|
|
|
Db 241 GTCCTCTTGAAGGGGACAGACCACTTCTCACTTAATCTTGGGTGTACCTTCCACTATA 300
|
|
|
QY 601 AACACCGCAAAATTAATGATCTTTCAGAGCCATGCAACTTGTATCGAGCTAACCAAAA 660
|
|
|
Db 301 AACACCGCAAAATTAATGATCTTTCAGAGCCATGCAACTTGTATCGAGCTAACCAAAA 360
|
|
|
QY 661 GAGCCTTTGTATCATAAATTTGATTACGATTCGAATTTTCAACGAAGAAATTTGATCAATCA 720
|
|
|
Db 361 GAGCCTTTGTATCATAAATTTGATTACGATTTTCAACGAAGAAATTTGATCAATCAATAGG 420
|
|
|
QY 721 GGACAAGCCTTGGATATATATCTGGAGAGACTTTTGCCTGAAATCATACCTACTCAGGAG 780
|
|
|
Db 421 GGACAAGCCTTGGATATATATCTGGAGAGACTTTTGCCTGAAATCATACCTACTCAGGAG 480
|
|
|
QY 781 ATGTATTGAATATGTTATGAATAAAGAGGGGAGTATCTTTTCCCATCGTCCAC 1020
|
|
|
Db 481 ATGTATTGAATATGTTATGAATAAAGAGGGGAGTATCTTTTCCCATCGTCCAC 720
|
|
|
QY 1021 GCCCTTAATCTTCACTAAAGAGAAAGGTCAAACTGAGCAACAATGAAATCTTAAGAAT 1080
|
|
|
Db 721 GCCCTTAATCTTCACTAAAGAGAAAGGTCAAACTGAGCAACAATGAAATCTTAAGAAT 780
|
|
|
QY 1081 CTCCTGTTGAGGACAACTGATTAAGATATAAAGTAAAGCTGATTCAAATGAGTGAATTC 1140
|
|
|
Db 781 CTCCTGTTGAGGACAACTGATTAAGATATAAAGTAAAGCTGATTCAAATGAGTGAATTC 840
|
|
|
QY 1141 GACACCAATTCATTTGGCTTACACCAAAAATTTTATTAATCAATTAAGTGAATATAAATA 1200
|
|
|
Db 841 GACACCAATTCATTTGGCTTACACCAAAAATTTTATTAATCAATTAAGTGAATATAAATA 900
|
|
|
QY 1201 AATGATATGAATAAGTATTTACTGATTTGGCTTCGATTCGACACCGCCACCAAT 1260
|
|
|
Db 901 AATGATATGAATAAGTATTTACTGATTTGGCTTCGATTCGACACCGCCACCAAT 960
|
|
|
QY 1261 TTACATGACGAATTTGTTATATATATAGACCACTTATCCGAATGTGA 1308
|
|
|
Db 961 TTACATGACGAATTTGTTATATATATAGACCACTTATCCGAATGTGA 1008
|
|
|
RESULT 7
```

AY692852	AY692852	1008 bp	DNA	linear	PLN 11-AUG-2004
LOCUS	Saccharomyces cerevisiae clone FLH13088.01X YPL069C gene, complete cds.				
DEFINITION	AY692852				
ACCESSION	AY692852.1	GI:51013154			
VERSION	Yeast ORF Project.				
KEYWORDS	Saccharomyces cerevisiae (baker's yeast)				
SOURCE	Saccharomyces cerevisiae				
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
REFERENCE	1 (bases 1 to 1008)				
AUTHORS	Marsischky, G., Rolfs, A., Richardson, A., Kane, M., Baqui, M., Taycher, E., Hu, Y., Vamberg, F., Weger, J., Kramer, J., Moreira, D., Kelley, F., Zuo, D., Raphael, J., Hogle, C., Jepson, D., Williamson, J., Camargo, A., Gonzaga, L., Vasconcelos, A.T., Simpson, A., Kolodner, R., Harlow, E. and LaBaer, J.				
TITLE	Creation of the YFLEX clone resource: cloning of Saccharomyces cerevisiae ORFs in the Gateway recombinational cloning system				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 1008)				
AUTHORS	Marsischky, G., Rolfs, A., Richardson, A., Kane, M., Baqui, M., Taycher, E., Hu, Y., Vamberg, F., Weger, J., Kramer, J., Moreira, D., Kelley, F., Zuo, D., Raphael, J., Hogle, C., Jepson, D., Williamson, J., Camargo, A., Gonzaga, L., Vasconcelos, A.T., Simpson, A., Kolodner, R., Harlow, E. and LaBaer, J.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-JUL-2004) Biological Chemistry and Molecular Pharmacology, Harvard Institute of Proteomics, 320 Charles st., Cambridge, MA 02141, USA				
COMMENT	This clone is part of a collection of Saccharomyces cerevisiae full-length ORF clones generated by the Harvard Institute of Proteomics. Each CDS has been cloned with its native stop-codon. The CDS has been directionally cloned using the Gateway cloning system into the donor vectors pDONR 201 or pDONR 221. Additional sequences in the clone: 'TCCAGTGCACACC' after the attL1 site and before the 'ATG' (from Research Genetics primers used to amplify the ORFs, including a Kozak consensus sequence); 'ATCCCGGAAATGCCAG' after the stop codon and before the attL2 site (from the Research Genetics primers used to amplify the ORFs).				
FEATURES	<p>Location/Qualifiers</p> <p>1..1008</p> <p>/organism="Saccharomyces cerevisiae"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:4932"</p> <p>/clone="FLH13088.01X"</p> <p>/lab_host="DH5alpha T1 resistant"</p> <p><1..>1008</p> <p>/product="YPL069C"</p> <p>1..1008</p> <p>/codon_start=1</p> <p>/product="YPL069C"</p> <p>/protein_id="AAT92871.1"</p> <p>/db_xref="GI:51013155"</p> <p>translation="MEAKIDELINNDPWSQNESLTKFPYHLLKPKGNRLNLIIVQINRVMLPKDQIAIVSQIVVELLHNSLDDIEDNAPRRGOTTSHLIFGVSTINTANYMYPRAMQVLSQLTKEPYLHNLITIEELINLHGGGLDLYWRDFPEIIPQEQMYLNMVWNTKGGFLRLRLMEALSPSHHGLSVLPINLGLIYYQIRDDYLNKQPOMSERGFADITTEGLKSPFIVHALNFTKTKGQEQHNEIRILLRLTRSDKDKLKLQILFEDFTNSLAYTKNFQINLVNMIKNENKYLPLDASHSDPATNLDHLLYIIDHUSEL"</p>				
ORIGIN	<p>Query Match 64.2%; Score 1008; DB 8; Length 1008;</p> <p>Best Local Similarity 100.0%; Pred. No. 5.9e-195;</p> <p>Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>				
Qy	301	ATGAGGCCAGATAGTACGCTGATCAATATGATCTGTTGGTCCAGCCAAATGAA	360		
Db	1	ATGAGGCCAGATAGTACGCTGATCAATATGATCTGTTGGTCCAGCCAAATGAA	60		
Qy	361	AGCTTGATTTCAAAACCTTATAATCACATCTCTTTGAAACCTGGCAAGACTTTAGACTA	420		

Db	61	AGCTTGATTTCAAAACCTTATAATCACATCTCTTTGAAACCTGGCAAGACTTTAGACTA	120
Qy	421	RAATTAATAGTTCAAAATTAACAGAGTTATGAATTTGGCCAAAGACAGCTGGCCATAGTT	480
Db	121	RAATTAATAGTTCAAAATTAACAGAGTTATGAATTTGGCCAAAGACAGCTGGCCATAGTT	180
Qy	481	TCGCAAAATGTTGAGCTCTTGATATAATTCACGCTTTTAATTCAGATATAGAAGTAAT	540
Db	181	TCGCAAAATGTTGAGCTCTTGATATAATTCACGCTTTTAATTCAGATATAGAAGTAAT	240
Qy	541	GCTCCCTTGAGAGGGGACAGACCTTCTCACTTAATCTTCGCTGTACCTCCACTATA	600
Db	241	GCTCCCTTGAGAGGGGACAGACCTTCTCACTTAATCTTCGCTGTACCTCCACTATA	300
Qy	601	AACACGGCAAAATATATATGATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCAAAA	660
Db	301	AACACGGCAAAATATATGATTTTCAGAGCCATGCAACTTGTATCGCAGCTAACCAAAA	360
Qy	661	GAGCCTTTGATCATATAATTTGATGATTTTCAAGGAAGATTTGATCAATCTACATAGG	720
Db	361	GAGCCTTTGATCATATAATTTGATGATTTTCAAGGAAGATTTGATCAATCTACATAGG	420
Qy	721	GGACAGGCTTGATATATGATTTTCAGAGCTTTCTGCCTGAAATCATACTACTCAGGAG	780
Db	421	GGACAGGCTTGATATATGATTTTCAGAGCTTTCTGCCTGAAATCATACTACTCAGGAG	480
Qy	781	ATGTATTTCATATATGTTGATTAATAAACAAGCGGCTTTTTCAGATTAACTGTGAGACTC	840
Db	481	ATGTATTTCATATATGTTGATTAATAAACAAGCGGCTTTTTCAGATTAACTGTGAGACTC	540
Qy	841	ATGGAAGCGCTGTCCTTCCTCACACCGGCCAATCGTTGGTTCCTTTTCAATAATCTT	900
Db	541	ATGGAAGCGCTGTCCTTCCTCACACCGGCCAATCGTTGGTTCCTTTTCAATAATCTT	600
Qy	901	CTGGGTATTTATTCAGATTAGAGATGATTAATTCGAATTTTGAAGATTTTCCAAATGTC	960
Db	601	CTGGGTATTTATTCAGATTAGAGATGATTAATTCGAATTTTGAAGATTTTCCAAATGTC	660
Qy	961	AGCGAAAAGGCTTTGCTCAGAGCATTACAGAGGGGAAGTTATCTTTTCCATCGTCCAC	1020
Db	661	AGCGAAAAGGCTTTGCTCAGAGCATTACAGAGGGGAAGTTATCTTTTCCATCGTCCAC	720
Qy	1021	GCCTTTAACTTCACTAAAACGAAAGTCAAACTGAGCAACAATGAAATTTCTAAGAATT	1080
Db	721	GCCTTTAACTTCACTAAAACGAAAGTCAAACTGAGCAACAATGAAATTTCTAAGAATT	780
Qy	1081	CTCCTGTTGAGCAAGTGTATAAGATATAAACTAAAGCTGATTCAAATCTGGAATTC	1140
Db	781	CTCCTGTTGAGCAAGTGTATAAGATATAAACTAAAGCTGATTCAAATCTGGAATTC	840
Qy	1141	GACACCAATTCATTGGCCTACACCAAAATTTTATTAAATCAATAGTGAATATGATAAAA	1200
Db	841	GACACCAATTCATTGGCCTACACCAAAATTTTATTAAATCAATAGTGAATATGATAAAA	900
Qy	1201	AATGATATGAAAATAGTATTTACCTGATTTGGCTTCGATTCGACACCGCCACCAAT	1260
Db	901	AATGATATGAAAATAGTATTTACCTGATTTGGCTTCGATTCGACACCGCCACCAAT	960
Qy	1261	TTATCATGACCAATTTGTTTATATATAATAGACCACTTATCCGAATTTGCA 1308	
Db	961	TTATCATGACCAATTTGTTTATATATAATAGACCACTTATCCGAATTTGCA 1008	

RESULT 8	AR071896	Sequence 1 from patent US 5912154.	1005 bp	DNA	linear	PAT 18-FEB-2000
LOCUS	AR071896					
DEFINITION	AR071896					
ACCESSION	AR071896					
VERSION	AR071896.1	GI:7222784				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unclassified.					

REFERENCE 1 (bases 1 to 1005)
Ferro-Novick, S. and Jiang, Y.
AUTHORS Geranylgeranyl diphosphate synthase proteins, nucleic acid
TITLE molecules and uses thereof
JOURNAL Patent: US 5912154-A 1 15-JUN-1999;
FEATURES Location/Qualifiers
source i.1005
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 64.1%; Score 1005; DB 6; Length 1005;
Best Local Similarity 100.0%; Pred. No. 2.4e-194;
Matches 1005; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 ATGGAGGCCAAGATAGATGAGCTGATCAATATGATCCCTGTTGTCCTCAGCCCAAAATGAA 360
DB 1 ATGGAGGCCAAGATAGATGAGCTGATCAATATGATCCCTGTTGTCCTCAGCCCAAAATGAA 60
QY 361 AGCTTGATTTCAAAACCTTTATATACATCCCTTTTGAACCTGGCAAGAACTTTAGACTA 420
DB 61 AGCTTGATTTCAAAACCTTTATATACATCCCTTTTGAACCTGGCAAGAACTTTAGACTA 120
QY 421 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATAGTT 480
DB 121 AATTTAATAGTTCAAAATTAACAGAGTTATGAATTTGCCCAAGACAGCTGGCCATAGTT 180
QY 481 TCGCAAAATGTTGAGCTCTTGATATATCCAGCCCTTTTAAATCGAGATATAGAAATAT 540
DB 181 TCGCAAAATGTTGAGCTCTTGATATATCCAGCCCTTTTAAATCGAGATATAGAAATAT 240
QY 541 GTCCTCTTGAAGGGGAGAGCCACTTCTCACTTAATCTTCGGTGTACCCCTCCACTATA 600
DB 241 GTCCTCTTGAAGGGGAGAGCCACTTCTCACTTAATCTTCGGTGTACCCCTCCACTATA 300
QY 601 AACACCGCAAAATTAATGATTTTCAAGCCATGCACTTTGATTCGAGCTAACCAAAAA 660
DB 301 AACACCGCAAAATTAATGATTTTCAAGCCATGCACTTTGATTCGAGCTAACCAAAAA 360
QY 661 GAGCCTTTGATTCATAATTTGATTAAGATTTTCAAGCCATGCACTTTGATTCGAGCTAACCAAAAA 720
DB 361 GAGCCTTTGATTCATAATTTGATTAAGATTTTCAAGCCATGCACTTTGATTCGAGCTAACCAAAAA 420
QY 721 GGAAGAAGCTTGATATATATCTGAGAGACTTTCTGCTGAAATCAATCACTACTCAGAG 780
DB 421 GGAAGAAGCTTGATATATATCTGAGAGACTTTCTGCTGAAATCAATCACTACTCAGAG 480
QY 781 ATGATTTGAATATGTTTATCAATAAACAAGCGGCTTTTTCAGATTAACGTTGAGACTC 840
DB 481 ATGATTTGAATATGTTTATCAATAAACAAGCGGCTTTTTCAGATTAACGTTGAGACTC 540
QY 841 ATGGAAGCGCTGCTCTCTCTCTCACACCGGCAATTCGTTGTTTCTTTTCATAAATCTT 900
DB 541 ATGGAAGCGCTGCTCTCTCTCTCACACCGGCAATTCGTTGTTTCTTTTCATAAATCTT 600
QY 901 CTGGGTATATTTACAGATTTAGATGATTTTGAATTTGAAGATTTTCCAAATGTC 960
DB 601 CTGGGTATATTTACAGATTTAGATGATTTTGAATTTGAAGATTTTCCAAATGTC 660
QY 961 AGCGAAAAGGCTTTGCTGAGGACATTTACAGAGGGAGTTTATCTTTTCCATCGTCCAC 1020
DB 661 AGCGAAAAGGCTTTGCTGAGGACATTTACAGAGGGAGTTTATCTTTTCCATCGTCCAC 720
QY 1021 GCCCTTAATCTTCACTAAACCAAGGTTCAAACTGAGCAACAATGAAATTTCTAAGATTT 1080
DB 721 GCCCTTAATCTTCACTAAACCAAGGTTCAAACTGAGCAACAATGAAATTTCTAAGATTT 780
QY 1081 CTCCTGTGAGGACAGGTGATAAGATATAAACTAAAGCTGATTTCAAAATCTGGAATTC 1140
DB 781 CTCCTGTGAGGACAGGTGATAAGATATAAACTAAAGCTGATTTCAAAATCTGGAATTC 840
QY 1141 GACACCAATTCATTGGCTTACACCAAAATTTTAAATCAATTAAGTGAATGATGATAAA 1200

841 GACACCAATTCATTGGCCTACACCAAAATTTTATTAATCAATTAAGTGAATGATAAA 900
QY 1201 AATGATATGAAAAATAAGTATTACTGTTGGCTTGCATTTCCGACACCGCCACCAAT 1260
DB 901 AATGATATGAAAAATAAGTATTACTGTTGGCTTGCATTTCCGACACCGCCACCAAT 960
QY 1261 TTACATGAGGAATTTGTTATATATATAGACCACTTATCCGAATTG 1305
DB 961 TTACATGAGGAATTTGTTATATATATAGACCACTTATCCGAATTG 1005

RESULT 9
CR380954.04
WPCOMMENT

Sequence split into 11 fragments LOCUS CR380954 Accession CR380954

Fragment Name Begin End
CR380954_00 1 110000
CR380954_01 100001 210000
CR380954_02 200001 310000
CR380954_03 300001 410000
CR380954_04 400001 510000
CR380954_05 500001 610000
CR380954_06 600001 710000
CR380954_07 700001 810000
CR380954_08 800001 910000
CR380954_09 900001 1010000
CR380954_10 1000001 1050361

Continuation (5 of 11) of CR380954 from base 400001 (CR380954 Candida glabrata strain CB8)

Query Match 17.7%; Score 277.2; DB 8; Length 110000;
Best Local Similarity 53.8%; Pred. No. 1.6e-46;
Matches 523; Conservative 0; Mismatches 523; Indels 12; Gaps 2;

QY 337 CCTGTTTGGTCCAGCCAAATGAAAGCTTGATTTCAAACCTTATAATCAATCAATGCA 396
DB 108599 CCAGCATGGTCACAATCAGATGATAAATTCATAAATCAACCATCAATCAATGCA 108558
QY 397 AAACCTGCGAAGAACTTTAGACTAAATTTAATAGTTCAAATTAACAGAGTTATGAA 456
DB 108559 AAACCGGGAAGAACTTTAGACTAAATTTAATAGTTCAAATTAACAGAGTTATG 108718
QY 457 CCCAAAGACCAAGCTGGCCATAGTTTGCATAATGTTGAGCTTTCGATAATTCAGCC 516
DB 108719 GATGAGATATATACAAACACTCACTCAGATTTGAGAGATCTCCCAATTTCAAG 108778
QY 517 TTAATCGAGATATAGAGATAATGCTCCCTTGAGAGGGGACAGACCACTTCTCATT 576
DB 108779 ATTATCGATGACATTTGAAGATACTCACAGTTGAGGGGTGAGTAGTCGCAATG 108838
QY 577 ATCTTCGGGTACCCTCCACTATATAACACCGCAATTAATGATTTTTCAGAGCATG 636
DB 108839 CTCTATGGTGTACCCTATGTCATTAATTAATGATTTTTCAGAGCATGATGAT 108898
QY 637 CTCTGATCGAGCTTAACACCAAGAGCCCTTTGATCATATAATTTGATTTACGATTT 696
DB 108899 TTACTTAGAAGCTGCTCATGATGAGTAACTTCAATGATCTAATGAAATTAATTA 108958
QY 697 GAAGAAATGATCAATCTACATAGGGCAAGGGCTTGGATATATATCTGGAGAGACTTT 756
DB 108959 GAGGAACCTTCTATATCTTACAGAGGCCAAGGATTTGATATATATCTGGCGTACT 109018
QY 757 CCTGAAATCATACTACTCAGGAGATGATTTTGAATGTTTGAATGATTAATAAACAAG 816
DB 109019 CCTAAGGTTGTACCAAGAGGAATGATTTTCAACATGTTTATGATAAATCTG 109078
QY 817 CTCTTTCAGATTAACGTTGAGACTCATGGAAGCGCTGCTCTCTCCCTCACACCGGC 876
DB 109079 CTCTTTCAGATTAACGTTTAAATTCATGGAAGCGCTTAAGCGAATCTGGCAG 109138
QY 877 TCGTTGGTTCCTTTCAATAATCTTCTGGGTATTTATTTATCAGATTTAGATGATTA 936
DB 109139 TCATTAGTTCCATTGGGAATTTTACTAGGAATTTATATATCAAGTTAGAGACGATTA 109198

QY	937	AATTGAAAGATTTC	996	Db	8719	GATGAGAATATTAT	8776
Db	109199	AATCTTACTGACTC	109258	QY	517	TTAATCGACGATAT	576
QY	997	AAGTTTATCTTTT	1056	Db	8779	ATTATCGATGACAT	8838
Db	109259	AAATTAATCTTTT	109312	QY	577	ATCTTCGGTGTGAT	636
QY	1057	CAACAATGAAATTC	1116	Db	8839	CTCTATGGTGTGAT	8898
Db	109313	AACAACACATCTG	109372	QY	637	CTTGATGCGACGTA	696
QY	1117	AGCTGATTCGAAT	1176	Db	8899	TTACTTAGAAACCT	8958
Db	109373	AATGCCATACGAT	109432	QY	697	GAAGAAATGATCA	756
QY	1177	AATCAATAGTGAAT	1236	Db	8959	GAGGAACCTTCTAT	9018
Db	109433	AAAAAATTCAGAC	109492	QY	757	CCTGAAATCATAC	816
QY	1237	TGCAATTCGACAC	1296	Db	9019	CCTAAGGTTGTAT	9078
Db	109493	ACCAATAGCCAC	109546	QY	817	CTTTTCAGATTA	876
QY	1297	TCCGAATTCGAAT	1356	Db	9079	CTCTTCAGATTA	9138
Db	109547	GCTAATGTGTGAT	109606	QY	877	TCGTTGGTTCCCT	936
QY	1357	CCTTCTCTCTCT	1416	Db	9139	TCATTAAGTTCC	9198
Db	109607	CCTGACCTTCT	109656	QY	937	AAATTTGAAAGAT	996
QY	1417	ATCTTTTGATAT	1476	Db	9199	AAATCTTACTGAC	9258
Db	109667	TTGATGATTTCT	109726	QY	997	AGTTATCTTTTCC	1056
QY	1477	TAGCCTTTATTT	1494	Db	9259	AAATTTATCTTT	9312
Db	109727	CATAAATAATGTT	109744	QY	1057	CAACACAATGAA	1116
RESULT 10				Db	9313	AACAACACATGCT	9372
CR380954_05				QY	1117	AAGCTGATTCAA	1176
WPCOMMENT				Db	9373	AATGCCATACG	9432
Sequence split into 11 fragments				QY	1177	AATCAATTAGT	1236
LOCUS CR380954				Db	9433	AAAAAATTCAG	9492
Accession CR380954				QY	1237	TCGCATTTCCG	1296
Fragment Name	Begin	End		Db	9493	ACCAATAGCC	9546
CR380954_00	1	110000		QY	1297	TCCGAATTCGA	1356
CR380954_01	100001	210000		Db	9547	GCTAATGTGT	9606
CR380954_02	200001	310000		QY	1357	CCTTCTCTCT	1416
CR380954_03	300001	410000		Db	9607	CCTGACCTTCT	9666
CR380954_04	400001	510000		QY	1417	ATCTTTTGAT	1476
CR380954_05	500001	610000		Db	9667	TTGATGATTT	9726
CR380954_06	600001	710000		QY	1477	TAGCCTTTAT	1494
CR380954_07	700001	810000		Db	109727	CATAAATAAT	109744
CR380954_08	800001	910000		QY			
CR380954_09	900001	1010000		Db			
CR380954_10	1000001	1050361		QY			
Continuation (6 of 11) of CR380954				Db			
from base 500001 (CR380954				QY			
Candida glabrata strain CE				Db			
Query Match	17.7%	Score 277.2;		QY			
Best Local Similarity	53.8%;	DB 8;		Db			
Mismatches	623;	Conservative		QY			
Indels	12;	Gaps		Db			
337	CCTGTTTGGTCC	396		QY			
8599	CCAGCATGGTCA	8658		Db			
397	AAACCTGCGA	456		QY			
8659	AACCCGGGA	8718		Db			
457	CCCAAGACCA	516		QY			

ETIT

Db	35688	TTTTTCCACGATTCTTGAACCTCTTTTATCAGCACCGTTAATGCTAGCGGTACTGTC	35622
QY	1549	AAATCGCCGGTAAATTCGCGA	1569
Db	35628	AAATCGCCGGTAAATTCGCGA	35608
RESULT 12			
LOCUS	AE016894/c	300316 bp DNA linear	PLN 14-APR-2004
DEFINITION	Eremothecium gossypii chromosome V, section 1 of 5 of the complete sequence.		
ACCESSION	AE016894		
VERSION	AE016894.1	GI:44983156	
KEYWORDS	Eremothecium gossypii (Ashbya gossypii)		
SOURCE	Eremothecium gossypii		
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Eremothecium.		
REFERENCE	1 (bases 1 to 300316)		
AUTHORS	Dietrich, F.S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., Mohr, C., Pohlmann, R., Luedi, P., Choi, S., Wing, R.A., Flavler, A., Gaffney, T.D. and Philippsen, P.		
TITLE	The Ashbya gossypii genome as a tool for mapping the ancient Saccharomyces cerevisiae genome		
JOURNAL	Science 304 (5668), 304-307 (2004)		
PUBMED	15001715		
REFERENCE	2 (bases 1 to 300316)		
AUTHORS	Gates, K., Dietrich, F.S., Brachat, S., Voegeli, S.E., Lerch, A., Philippsen, P. and Gaffney, T.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-DEC-2002) Department of Ashbya Genomics, Syngenta, Research Triangle Park, NC 27709, USA		
COMMENT	This is low coverage sequence generated to identify the complete set of genes and the gene order on this chromosome. Regions of low quality are identified. Before doing extensive work on any gene identified here the sequence should be confirmed.		
FEATURES	Location/Qualifiers		
source	1..300316		
	/organism="Eremothecium gossypii"		
	/mol_type="genomic DNA"		
	/strain="ATCC 10895"		
	/db_xref="ATCC:10895"		
	/db_xref="taxon:33169"		
	/chromosomes="V"		
source	/note="synonym: Ashbya gossypii"		
	48858..92110		
	/organism="Eremothecium gossypii"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:33169"		
	/clone="BAC clone BAG1539"		
source	62601..127208		
	/organism="Eremothecium gossypii"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:33169"		
	/clone="BAC clone BAG1581"		
source	103331..181613		
	/organism="Eremothecium gossypii"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:33169"		
	/clone="BAC clone BAG1047"		
source	169336..224489		
	/organism="Eremothecium gossypii"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:33169"		
	/clone="BAC clone BAG1442"		
source	210657..270634		
	/organism="Eremothecium gossypii"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:33169"		
	/clone="BAC clone BAG1767"		
source	233952..295486		
	/organism="Eremothecium gossypii"		

Db	189975	TGTTCAACGGGTTCTATGTTCTTTTCGAGCGGCGAGTTCGCGCGGTTCACAGAGTTGGTGG	189916		
QY	494	AGCTCTTGCAATAATCCAGCCCTTTTAATCGACGATATAGAGATAATGCTCCCTTGAGAA	553		
Db	189915	AGATGTTGATGTGGGAGCTGTTGATCGACGATATGAGACAACTCTGATGAGAC	189856		
QY	554	GGGGACAGACCACTTCTCACTTAATCTTCGGTGTACCTTCCACTATAAAACACGGAAATT	613		
Db	189855	GTGGCGTCTGCTGCCACGTTGTATATGGCTCGCCCATGACGATCAACACTTGCGAATT	189796		
QY	614	ATATGATTTTCAGAGCCATGCAACTTGTATGTCAGCTAAACCAAAA-----AGAGCCTT	667		
Db	189795	ACATGATTTTGTGTCATGCTGCTCGGCGAGCTGGCGGTACGCGTTCGCGCGGGC	189736		
QY	668	TGTATCATATTTGATTAACGAAATTTCAACGAGAAATGATCAATCTACATAGGGGCAAG	727		
Db	189735	CGCTGCAGGATCTGCTGAAGGTGTTCAACAGAGAGATGATGAATCTGCATCGCGGCGAG	189676		
QY	728	GCTTGGATATATCTCGAGAGACTTTCGCTGAAATCATCACTACTCAGGAGATGATTT	787		
Db	189675	GCCTTGACATATTTGGCGGACACGTTTACGG-----TGCGTGGAGCATGACTACC	189622		
QY	788	TGAATATGTTATGAATAAAACAGCGCGCTTTTCAGATTTAAGCTTGAGACTCATGGAAG	847		
Db	189621	TGCGGATGGTGAATGATAGACCGCGCGCTTGTTCGGCTGACCGTGGTATCATGGAAG	189562		
QY	848	CGCTGCTCTCTCTCTACA---CCAGCGCAATTCGTTGGTCTCTTCATATAAATCTTCTGG	904		
Db	189561	CGTTACGGGAGGCGCCAGATGGAGCCCGGATCAACATTTGTGCGCGTTAGCAACTGCTTG	189502		
QY	905	GTAATTTATTCAGATTAGAGATGATTTGATTTGAAAGATTTTCCAAATGTCAGCG	964		
Db	189501	GCCTGCTGTACCAAGTGGCGGACACTATTTGAACCTGACCGACTCGCGCATGTCGAAA	189442		
QY	965	AAAAAGGCTTTGCTGAGGACATTCACAGGGGAAAGTTATCTTTTCCCATCGTCCACGCC	1024		
Db	189441	ACAAGGCTTTGACAGCATATCACGAGGGGCAATTTCTCTACCCCATTTATCCACGGCC	189382		
QY	1025	TTAACTTCATAAAACGAAAGGTCAAACTGACCAACCAATGAATTTCTAAGAAATCTCC	1084		
Db	189381	TGCAGTACGCCCGTGTCCA---TGACCCAGCGGCTACGACTTCTCGTCAGTGTGTAC	189325		
QY	1085	TGTTGAGGACAAGTGATAAGATATAAACTAAGCTGATTTCAATATCTGGAATTCGACA	1144		
Db	189324	GGCAGAGACGACAGATATCACCAAGCGCGGTGGTGGGATATCTTGGCGAGCTCA	189265		
QY	1145	CCAATTCATTTGGCTACACCAAAATTTTATTAATCAATTAAGTGAATATGATAAAA	1201		
Db	189264	GTGGTTCGCTCGGTACACGAGCAGCGGTATCATCGAACTAGCAACATTGATAAAGA	189208		
RESULT 13	CNS06EHJ	952 bp	DNA	linear	STS 10-JAN-2001
LOCUS	T3 end of clone AROAA021F04 of library AROAA from strain CBS 732 of				
DEFINITION	Zygosaccharomyces rouxii, sequence tagged site.				
ACCESSION	AL395117				
VERSION	AL395117.1				GI:12146499
KEYWORDS	STS.				
SOURCE	Zygosaccharomyces rouxii				
ORGANISM	Zygosaccharomyces rouxii				
REFERENCE	Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
AUTHORS	Saccharomycetales; Saccharomycetaceae; Zygosaccharomyces.				
	1 (bases 1 to 952)				
	Souciot, J.L., Aigle, M., Artiguenave, F., Blandin, G.,				
	Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,				
	de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Liorens, B.,				
	Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,				
	Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,				
	Wincker, P. and Weissenbach, J.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of				
	yeast species for molecular evolution studies				
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)				
MEDLINE	20584711				
PUBMED	11152876				
REFERENCE	2 (bases 1 to 952)				
AUTHORS	de Montigny, J., Straub, M., Potier, S., Tekala, F., Dujon, B.,				
	Wincker, P., Artiguenave, F. and Souciot, J.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 8.				
JOURNAL	Zygosaccharomyces rouxii				
MEDLINE	FEBS Lett. 487 (1), 52-55 (2000)				
PUBMED	20584718				
REFERENCE	3 (bases 1 to 952)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,				
	2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :				
COMMENT	seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)				
	This STS is part of a random genomic sequencing program of thirteen				
	yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces				
	exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,				
	Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces				
	lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia				
	angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,				
	Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to				
	5 kb were prepared and both extremities were sequenced. See				
	keywords for description of this sequence and for the sequence of				
	the other extremity of this insert.				
FEATURES	Location/Qualifiers				
source	1. .952				
	/organism="Zygosaccharomyces rouxii"				
	/mol_type="genomic DNA"				
	/strain="CBS 732"				
	/db_xref="taxon:4956"				
	/clone="AROAA021F04"				
	/clone_lib="AROAA"				
	/note="end : T3"				
misc_feature	<154..>270				
	/note="similar to Saccharomyces cerevisiae ORF YPL070w [
	weak similarity to Vps9p]"				
	/evidence="not experimental"				
misc_feature	complement(<435..>908)				
	/note="similar to Saccharomyces cerevisiae ORF YPL069c [
	BTS1 ; geranylgeranyl diphosphate synthase]"				
	/evidence="not_experimental"				
ORIGIN					
Query Match	8.2%; Score 128.6; DB 11; Length 952;				
Best Local Similarity	56.6%; Pred. No. 5.6e-16;				
Matches	303; Conservative 1; Mismatches 215; Indels 16; Gaps 3;				
QY	679	TTGATTACGATTTTCAACGAGAAATGATCAATCTACATAGGGGACAAGCTTGGATATA	738		
Db	949	TTTAAATGTTTAYCTTCAATCAGGAAATGATTAAATTTGCATA-GAGCCAAAGGTTTAGATATT	891		
QY	739	TACTGGAGAGACTTTCGCGCTGAAATCATCTACTCAGGAGATGTTTGAATATGTT	798		
Db	890	TACTGGCGGATAATTTATTTCGAGTTTATACCGATGAGCGACGTACTATAATATGTTG	831		
QY	799	ATGAATAAAACAGCGCGCTTTTCAGATTAAACGTTTCAGACTCATGGAAGCGCTGTCCT	858		
Db	830	ATGAACAAGACTGCTGGCTTATTTAGATTAAACGGTTAGATCATGGAATGCTGACCGAT	771		
QY	859	TCCTCACACACCGCCCATTCGTTGGTTCTTTTCAATAAATCTTCTGGGTATTTATATCAG	918		
Db	770	---GTAGATTATCCGCATTTCATTAGTACCGCTGAGCAACCTTTTGGGAATAATTTATCAA	714		
QY	919	ATTAGAGATGATTTACTTGAATTTGAAAGATTTTCAAATGTCACGCGAAAGGCTTGTCT	978		
Db	713	ATTTCGCGATGATTTCAAAATCTTTTTCGAAATGCAAAATGATAGCTATAGAGGTTAGCA	654		
QY	979	GAGGACATTCACAGAGGGGAAGTATCTTTTCCCATCGTCCACGCGCTTAATCTCACTAAA	1038		
Db	653	GAAGATATTTCTGAGGTAACTATCTTTCCCGATTTATCCACGCGCTTGAAT-----	602		

Db	374	AGATTATCAATTGAAGTGTGCTGAAATGTTGTCATAATGCGCCAGTTTACTCATTTGATGATGATTG	433
Qy	533	AAGATAATGCTCCCTTGAGAAAGGGGACAGACACACTTCTCACTTAATCTTCGGTGTACCCCT	592
Db	434	AAGACAGTTTCAAGACTCGAGCTGGTTCAGTTTTCAGTTCACAGCAATCATGTGTGCCAT	493
Qy	593	CCACTATAAACACCGCAAAATATATGATTTTCAGAGCCATGCAACTGTGTATCGCAGCTAA	652
Db	494	CTGTCAITTAACCTGCGCAATATGTCTACTTCTTGGCGCTGGAAAAAGTCT	544
Qy	653	CCACAAAGAGCCCTTGTATCATTAATTTGATTAACGATTTTCAACGAGAATTGATCAATC	712
Db	545	-----TAACACTTGTATCACCCAGATGCGAGTGAAGCTTTTTCACAGCAGAGCTTTTGGAC	598
Qy	713	TACATAGGGGACAAGGCTTGGATATATACTCGAGAGACTTCTTGCTCGAAATCATACCTTA	772
Db	599	TTCAATCAGGACAGGCTAGATTAATTTACTGAGAGA-----TACCTFACATTTGTCTTA	652
Qy	773	CTCAGGAGATGATTTTGAATATGTTATGAATAAAACAGGGGGCCCTTTTCAGATTAACGT	832
Db	653	CTGAAGAAGAATAATAAAGCCATGTTTTCAGAGACAGGTGGCTTGTTCGATTAGCG	712
Qy	833	TGAGACTCATGGAAGCGCTGCTCTCTCCTCACACCGGCCATTCGTGTGTTTCCTTCA	892
Db	713	TAGGCCCTCATGCAGTTGTTCTCTGACTACAAAGAAGATCTAAAGCCACTGCTGG-----	766
Qy	893	TAAATCTCTGGGTATTATTTATCATGATTAGAGATGATTACTTGAAATTTGAAGATTTC	952
Db	767	---ACACACTTGGCTCTTCTTCAGATTAGAGATGATTACGCCAATCTACACTCCAAAG	823
Qy	953	AAATGTCCACGCAAAAGGCTTTGCTCAGGACATTACAGAGGGGAAAGTATCTTTTCCCA	1013
Db	824	AATACAGTGAATAAAGTTTCTGTGAAGACCTCACAGAAGGAAAGTCTCAITTCCTTA	883
Qy	1013	TCGTCCAGGCGCTT 1026	
Db	884	CTATTCATGCCATT 897	
RESULT	15		
AB118243			
LOCUS	AB118243	1487 bp	linear
DEFINITION	Rattus norvegicus rGPS1delta pseudogene mRNA.		ROD 21-JUL-2004
ACCESSION	AB118243		
VERSION	AB118243.1	GI:50428392	
KEYWORDS			
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
REFERENCE	1		
AUTHORS	Matsumura, Y., Kuzuguchi, T. and Sagami, H.		
TITLE	Relationship between Intron 4b Splicing of the Rat Geranylgeranyl		
JOURNAL	Diphosphate Synthase Gene and the Active Enzyme Expression Level		
REFERENCE	2. (bases 1 to 1487)		
AUTHORS	Matsumura, Y., Kuzuguchi, T. and Sagami, H.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-AUG-2003) Hiroshi Sagami, INRAM, Tohoku University;		
2-1-1 Katakura, Aoba, Sendai, Miyagi 980-8577, Japan			
(E-mail: yasagami@agen.tohoku.ac.jp, Tel: 81-22-217-5622,			
Fax: 81-22-217-5620)			
FEATURES			
source	Location/Qualifiers		
	1. .1487		
	/organism="Rattus norvegicus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10116"		
	/sex="male"		
	/tissue_type="testis"		
gene	1. .1487		
	/gene="rGPS1delta"		

